新 編

生物統計學

包興林 著

新竹黎明書店經銷

新編

生物統計學

包興林 著

新竹黎明書店經銷

目　　錄

第一章
統計學係搜集分析與整理統計資料並推論其結果的科學

一、一般統計資料的處理應保留多少位數合宜須加斟酌

幾乎世界上所有的資料，均須符合下述四種特點，方可叫做統計資料：

①數量性，所謂的數量性，即是可以數字來表現它特點的資料，如果是品質資料，則必須事先轉變爲數量資料之後再加以整理。

②眞實性，資料的統計，必須根據測量或登記、調查、實驗獲之，來自眞確而客觀之事實，並不是個人主觀意志來決定。

③變異性，統計上所探討的課題，絕大多數屬於變異的問題。如果一份資料中各方面的特性量的特徵或性質的特徵完全相符時，則此份資料極爲單純，即不必利用統計方法探究之。

④群體性，統計最重要的，是在研討群體的數量特徵，而不是零散的個別情事。譬如觀察中華民國的婦女，在生育期所生育的子女總數，而不是以特定的某一婦女或某一城市的婦女而言，而是必須囊括台澎金馬所有地區爲範圍的。

一般所稱之統計資料，是根據個體特性以點計（數數）或度量（測驗）自然景象、生物景象或社會景象各方獲得之群體的資料。這種統計資料，必須含有時間、空間與質量三個要素，群體內個體要素的多寡，和群體的恰成相反，群體中個體的要素多，群體的範圍也就愈小，反之

則大。

以生物現象的「人」來說，倘若僅根據「人」生理上必備的質的要素為主，（例如靈活運作的雙手、直立行走的特徵、以及胎生等等），而沒有參雜時間、空間以及量等要素為範疇，如此，地球上自古至今所有的人類，便都屬於人的群體。如果參與時間、空間以及量三大要素及其他更多層面的條件，則其群體的局限就將會狹小許多了。

統計學可分為應用統計學和純理統計學兩種，應用統計學是探討統計的方法於各種現實問題的處理上的有效運用，純理統計學則探討統計方法的原理，在本書中討論的生命、人口、醫學以及生物統計學，都是應用統計學的一部分。

我們所說的統計學，就是採集、整理和統計資料的分析，並且推論與詮釋最後結果的一種工具科學。依據我國所謂的統計學這個名詞，乃是從英文「Statistics」直接翻譯過來，這個字的德文拼音是「Statistik」、法文是「Statistique」，它們的讀音都頗為近似。直接從意大利文「Statisti」此字轉變而成，原本的意思是政治家「Statesman」；間接的起源是依據拉丁文「Status」而得之。「Status」原來的意思，是說明一國的政治情況（A political state）。然而，政治情況主要的也就是人口和財富，尤其是以人口為最重要的現象，人口是一國情勢表徵，國家為了需要徵兵，理所當然要詳加調查及登記這國家的武力根源——人口，來作為計劃兵源的根本依據。

最早的統計是起源於政治的需要而對人口大概的統計而來的。

人類的知識，難以千萬計，粗略的區分，大致可分為三種。一是經驗的知識，必須要長時間的親身體驗、經歷方可獲得，當時的人們往往只知其然而不知其所以然；第二種是科學的知識，不僅知其然，並也知其所以然，是經過專門的研究探討和反覆的實際應証或客觀的查証而獲之。就如同當今進步的文明社會中衆多機械工程、電、力、聲光與衞生

醫學各層各面的原理探究及發明，都是歸屬科學的知識。

第三種卽是逾越經驗知識與科學知識的統計知識，譬如生命現象中的「巨數恒性」的知識、生物在遺傳上顯性與隱性比例關係的知識、醫學治療上各個年齡以及不同性別的各項疾病發生率的知識，都是屬於這方面的知識。如果人類的知識中，缺乏了統計知識，則相信必有許許多多難以理解或茫然無知的情形發生。

天地之間的各種人、事、物，都是參差不齊、變化多端、難以捉摸，甚至時而消失、忽又出現。無論它如何的變異，只要仔細觀察，統計、研究，仍是有跡可尋的，這就是統計現象在一個龐大群體時，多是處於常態分配的情況，就拿人類的體質來說，人們的高矮、胖瘦、身強、體弱，普通是以中庸者爲多數，而愈向兩方極端則愈少。統計人員就依照此種事實，可用一部分來代替一個龐大群體，來攫取所需的代表數值，來應用於同一種類的事物。醫學與生物的研究實驗上，大部分應用統計現象的常態分配來整理資料。

至於科學的分類，各個科學家都有他們不同的獨到分法，例如孔德（ Comte ）將科學分爲化學、生物學、社會學、天文學、物理學、數學等六類，斯賓塞則將之分爲抽象與具體科學。大致來說，應分爲下述三類較適當，第一類是研討物質的無機現象的物質科學，也有人將之稱爲自然科學，包括物理學、化學、天文學、冶金學、工程學、建築學、地質學等均屬之；第二類是研究有機的生物現象的生物科學，此外，也有些學者將物質科學與生物科學併稱爲自然科學，有人口學、生物學、動物學、植物學、心理學、生理學、醫學、育種學、優生學等。第三類是研討超機的社會現象的社會科學，如社會學、政治學、經濟學、法理學、倫理學、商學、教育學等。

任何一類的科學都是需要用到統計方法，才能獲得越發精密且有價值的結果；社會科學的日精月益，絕大多是應用統計方法而得，社會現

象是雜亂無章的，必須藉著統計科學來整理，含混不清的事態也要靠著統計來分析，錯綜複雜的因果關係的問題也須利用統計來解答，奧秘難測的趨向亦是憑藉著統計來推斷，這種科學的利用價值眞可謂是博大精深了。

有句話說「數字是不會撒謊的」（Figures won't lie），可堪稱眞理。

以簡單的數字表示極其繁複的狀況，是統計法的優點。

統計的資料，必定是數量的資料。統計的結果，也必定是以數字來表示。統計上所有的數量，是由點計或者測度而得，不容易達到絕對的準確，僅是準確到足以取信之程度罷了。實際上，凡是經由測度而來的數量，姑且不管它是自然現象、生物現象或者是社會現象，要達到絕對的準確根本是不可能的，例如人體的脈搏每分鐘跳動七十二次、體溫多少、高低血壓各是多少等都是。點計所得到的數字或許眞能夠達到百分之百的絕對準確，但自點計數字演化而來的其他數字，如男女性別的比例、出生及死亡的比率等，都只是近似值，不可能成爲絕對正確值。如此，便產生了誤差。

通常統計資料的整理，保留多少位數方爲合宜，必須詳加研究。統計數値的計算，如果保留數字過少，會影響準確性，過多，又會加增計算上的繁複與混淆不清。一般均考慮到二方面來決定準確的單位，一：素材本身的特性，準確性的容許程度爲何？二：我們所需要的準確度爲何？例如平常算計長度，道路就以公里或市里計算，布料以尺或寸計算，身高論公分。至於重量、體重以公斤計算，黃金以兩或錢或分計算……等。基於事實的需要，不同的情況便要保留不同的數值，所以經常會删除或者簡化尾數。通常是用四捨五入法，但是這種簡化法，入的機率爲$\frac{5}{9}$，捨的機會爲$\frac{4}{9}$，似乎缺乏精準，因此，便有人提議，比五大的入，比五小的捨，剛好爲五的就看前一位數爲奇數時加一，前一位數爲偶

數時就捨去不要了，使其入捨機會均等。

誤差

①可能誤差與相對誤差

統計數值最末一位數的半個單位即爲可能誤差，譬如說一個人的身高是178公分，則可能誤差就是0.5公分，體重50.5公斤，那麼，可能的誤差就是0.05公斤。相對誤差爲可能誤差與近似值（統計數值）之比，以公式表示：

$$相對誤差=\frac{可能誤差}{統計數值}\times\frac{100}{100}$$

$$例如上述之身高的相對誤差=\frac{0.5}{178}\times\frac{100}{100}=0.28\%$$

$$體重的相對誤差=\frac{0.05}{50.5}\times\frac{100}{100}=0.1\%$$

誤差可能有單位，這個單位與原來統計數值的單位一樣，各統計數值誤差的大小，較不適宜用可能誤差來比較，以相對誤差比較較適當，原因是，相對誤差沒有單位，誤差的程度可明白的顯示出來。誤差的另一面則爲確度，所以相對誤差小，其準確度較高，相對誤差大，準確度便要低了，由此可得一結論，即爲，一數值的相對誤差和準確度之間的關係是互爲相反的。

②絕對誤差與實際誤差

所謂的絕對誤差就是，近似值與眞值之差，而實際誤差就是絕對誤差和眞值的比了。因爲眞值不容易求得，所以絕對誤差和實際誤差也就不好算出，通常即以可能誤差和相對誤差來替代。

所謂的有效數字，是指用來表示一數值的準確程度的數字。

通常計算有效數字的方法，是由數字的左邊數到右邊，而不是由零的第一位數字向左算到最後一位準確數字爲止，通常單純整數最末的零

不能算為有效數字，而小數末尾的零就必須算是有效數字了。

相對誤差的大小與有效數字位數的多少之間的關係是成相反的關係，有效數字位數多者相對誤差小，反過來說，有效數字位數少的話，相對誤差便大；有效數字位數若是相同，那麼數值中數字小的相對誤差大，數值中數字大的相對誤差就小。

此外，對於有效數字位數的計算方法，統計數值若是沒有整數，則小數點後面的零就不算作是所謂的有效數字了。

近似值的計算法

①近似值差數的兩個可能誤差，等於兩近似值的可能誤差之和。因為減法是加法的還原，所以減法準確度的計算和加法的計算並無軒輊，也是以有效數字位數最小的一個準確度而決定差數的準確度。近似值的和與差的算法也就是近似值的和數和差數的準確度的決定法。胡元璋在他所著的「統計計算的確度」一書中引述了金氏（ W. I. King ）的一段話：「一鍊的強度決定於最後的一環，一個總和的確度亦由其確度最低之一項決定之。」實在是精簡明瞭。

②幾個近似值和數的可能誤差，等於各近似值的可能誤差之和。

設有近似值 x_1 ，x_2 ，x_3 ……x_n 其可能誤差分別為 $\pm e_1$ ，$\pm e_2$ ，$\pm e_3$ ，……$\pm e_n$ ，則 x_1 之眞值應為 $x_1 \pm e_1$ ，……x_n 的眞值應為 $x_n \pm e_n$ 。設 x_1 ，x_2 ，……x_n 之總合為 x ，x 的可能誤差為 E ，則可以得到下列二式：

$$x = x_1 + x_2 + \cdots\cdots + x_n \qquad ⓐ$$

$$x \pm E = (x_1 \pm e_1) + (x_2 \pm e_2) \cdots\cdots + (x_n \pm e_n) \qquad ⓑ$$

ⓑ式－ⓐ式　$\pm E = \pm e_1 \pm e_2 \cdots\cdots \pm e_n$　此明顯的表示總和的可能誤差為原來各數值的可能誤差之和。

計算一準確度不同之近似值的和數時，應先把各數值化簡到相同的

單位時再相加，其和數的準確度以原來各數值中的有效數字位數最少的近似值的單位爲準。

商的準確度

乘法的還原是除法，所以除法計算的確度和乘法的是一樣的，除法的相對誤差，大抵等於被除數與除數之相對誤差的和，而商數的有效數字的位數，最多等於（一般情況是少於）兩近似值中有效數字的位數最少者。倘若相除的二個數，一爲精確值，一爲近似值，這個商數的有效數字位數，一般都是以近似值的有效數字位數爲準，如剛好可除盡，商數的有效數字位數少於其中近似值有效數字位數也可。

積的準確度

①積數的相對誤差，大抵等於相乘的二個數之相對誤差之和。

設 x_1，x_2 二數相乘，其相對誤差 e_1' 及 e_2'

則 x_1 之眞值爲 $x_1(1 \pm e_1')$

x_2 之眞值爲 $x_2(1 \pm e_2')$

二者相乘後之最大積數爲 $x_1(1+e_1')x_2(1+e_2')$，最小積數爲 $x_2(1-e_1')x_2(1-e_2')$。

其積數的相對誤差爲：

$$\frac{x_1(1+e_1')x_2(1+e_2')-x_1x_2}{x_1\ x_2}$$

$$=\frac{x_1x_2[(1+e_1')(1+e_2')-1]}{x_1\ x_2}$$

$$=(1+e_1')(1+e_2')-1$$

$$=1+e_1'+e_2'+e_1'\,e_2'-1$$

$$=e_1'+e_2'+e_1'\,e_2'$$

或爲

$$\frac{x_1x_2 - x_1(1-e_1')x_2(1-e_2')}{x_1\ x_2}$$

$$=\frac{x_1x_2\left[1-(1-e_1')(1-e_2')\right]}{x_1\ x_2}$$

$$=1-(1-e_1')(1-e_2')$$
$$=1-1+e_1'+e_2'-e_1'\ e_2'$$
$$=e_1'+e_2'-e_1'\ e_2'$$

以上左右二式中之$\pm e_1'\ e_2'$極為微小，為了方便計算，故以省略，如此左右二式均可改為：$e_1'+e_2'$，即積數之相對誤差為二數之相對誤差之和。故乘積的真值為：$x_1x_2\left[1\pm(e_1'+e_2')\right]$。

影響乘法最大的是準確度最低的一項，因此，積數的有效數字位數最多等於（或少於）兩近似值中有效數字之位數最少者。如果相乘的二個數，一為精確數，一為近似值，它的積數之有效數字位數，一般是以近似值的有效數字位數為準。

②積數的可能誤差，大抵等於乘數及被乘數和其可能誤差二者相乘之積的和。

設x_1，x_2二數相乘，其可能誤差為e_1'，e_2'，則積數的最大可能為：

$$(x_1+e_1)(x_2+e_2)$$
$$=x_1x_2+e_1e_2+(x_1e_2+x_2e_1)$$

積數的最小可能為：

$$(x_1-e_1)(x_2-e_2)$$
$$=x_1x_2+e_1e_2-(x_1e_2+x_2e_1)$$

故$(x_1\pm e_1)(x_2\pm e_2)$之積必為：

$$x_1x_2+e_1e_2\pm(x_1e_2+x_2e_1)$$

因e_1e_2之值甚小，為了方便計算，大多省略不計，故$(x_1\pm e_1)$

$(x_2 \pm e_2)$ 之積可改寫爲：

$$x_1 x_2 \pm (x_1 e_2 + x_2 e_1)$$

計算積數可能誤差的簡單方法可以下式代入：

$$(x_1 \pm e_1)(x_2 \pm e_2)$$
$$= x_1 x_2 \pm (x_1 e_2 + x_2 e_1)$$

偏誤與簡化尾數

上述之近似值的誤差，有正與負二個方向，設有近似值若干個，不一定誤差會相同，往往會有偏高或偏低者，計算後，可能正負號會互相抵銷，並非一定能達到誤差的極限，這叫作「非偏誤」（Unbiased error）或者是「補償誤差」（Compensative error）。倘若各數值都爲同號的誤差，則叫做「偏誤」（Biased error），或「累積誤差」（Cummulative error）。有一點須加以注意，原來的數值爲非偏誤時，和數與積數的誤差或許會減少，而商數與差數的誤差就不一定了。如果被除數及被減數偏高，而除數及減數偏低的話，商數和差數的誤差增大偏高，反之則增大偏低。再者，原來數質爲相同方向偏誤時，和數和積數增大偏誤，商數與差數便有相互抵銷而減輕偏觀的作用。這在四捨五入尾數的簡化也是一樣的，通常四捨五入後的數值，如果有偏誤，則會若干項偏高，若干項偏低，相乘或相加的誤差可降到最少。不過，在除法與減法中，若四捨五入後，一邊偏高一邊偏低的話，商數和差數的誤差就會增大。倘四捨五入後有相同方向偏誤，則商數和差數不變，但和數及積數便會增大差誤。所以，在四捨五入簡化尾數時，應該謹愼才好。

二、統計研究問題成級資料最爲簡便經濟

爲了解所有實際狀況，必須依據調查，如果將所要研討的全部實際

狀況詳加調查，以所有的資料作個分析與探討，則必爲最實在且可靠的資料。這種以某一特定情形進行總體個別調查的方法稱爲普查（ Census ），例如人口普查便是。可是，由於天地之間，無論何種現象，都是十分繁複，而且散佈又廣，往往會受人力、時間、經濟等各因素的限制，不可能作全面的調查，因此，除了特別重要的資料，很少應用到普查。即使是一國最重要的人口資料，也是要隔許多年方能舉行一次普查，並不是經常性的。

一般蒐集靜態原始資料，大部份是採用抽樣調查（ Sampling Survey ），意思就是，在全部的事物中抽取一部分，將所抽查的樣本（ Sample ）作爲全部事物的依據，以樣本特性來闡明總體的特性，如果抽樣得宜，則其正確度可達很高，如此，誤差便會降至最低程度。

抽樣的方法大致可區分如下：

①部落抽樣（ Cluster random sampling ）

此法與分類比例近似，都是先把待研究的事物依某項規則分爲若干類，每類就叫一部落，其次，自所有的部落中，隨便選幾個部落來進行調查，這種方法就叫作部落抽樣法。不過這要是各個部落間相差不懸殊，並且部落內部的分子差異大時比較準確。假使各個部落間差異懸殊，而部落之內部大致相仿，則這個部落抽樣就不適合使用了。

②隨意抽樣（ Random Sampling ）

另一名稱是「機會抽樣」，一個群體中的每一分子都有相同被列爲抽樣對象的機會，且不參雜研究人員的任何主觀意見。這種抽樣方式，是自所有全部的資料中挑選，而不是只在一部分資料中抽選，而且，所抽取的樣本範圍非常廣，通常是代表優質的。一般使用抽樣工具決定調查樣本。

③兩段抽樣（ Two stage sampling ）

首先，把待研究的事物分爲幾個組，用隨機法抽取其中幾組，其

次就在抽出的幾組中進行抽樣調查。

④分類比例抽樣(Stratified proportional random sampling)

也可叫作「分層比例抽樣」，是把待研究的事物依某種規格分爲幾個組，其次是在各組中，依照群體大小，個體多少的比例抽樣調查。

爲了節省人力、財力及時間，研究統計任何問題，若有現成次級資料可應用，則將簡便且經濟許多。次級資料的來源最主要的就是政府的行政機構所發表的調查統計。但在使用時應詳加研究是否符合使用的條件，不可盲目引用。

應用次級資料，應先就下述兩點加以審愼詳細考慮：

①次級資料的單位是什麼？前後是不是一致？是不是各地均相同？是否能符合所要研究調查的課題？次級資料的正確性是不是足夠？倘使同一事件同時有若干個次級資料的統計報告，應查其結論是否一致？假使有異，就要調查它不同的原因，同時，也要將各個不同的統計報告一起作個比較，再決定取捨。

②提供次級資料考查的組織或機構是什麼？是否曾有不名譽的歷史？它之所以提供資料的目的是什麼？這些種種的問題，都應在使用它的資料之前，先加以探究一番。此外，尙且須考慮到這個次級資料的調查方法是什麼？是普查抑或抽查？如果是抽樣調查，它抽樣的方式及樣本的形式是否符合我們的研究需要？

統計學資料的種類，依照資料曾否經過整理，可區分爲二：

①次級資料Secondary data：

是已經過第一手整理過、簡化過而轉錄的資料，已不是原始資料，所以也可說是間接資料。因爲這種資料可以立刻派上用場，因此也叫作現成資料Existing Information。

②原始資料Primary data：

又叫作直接資料，因爲它是統計人員直接取自來源處，尙未經過整理過程，保持著最初的狀態，所以也可稱之爲初級資料。

按照資料經歷的時間過程，可區分爲二：

①動態資料 Dynamic Data：

是指在一特定時期內演進更變的資料。動態資料之獲得大多是經由定期不斷的報表登記。

②靜態資料 Static Data：

是指在一特定時期內未發生演進更變而保持靜止狀態的資料而言。此靜態資料，大多是依據調查而得。

利用法令或規章加以限制，使動態資料的登記不會發生中斷的情形，促使制度化、正常化，應使用強制的手法使之成爲習慣性的必然工作，完成登記的表式、登記的機構及工作人員、登記的事項範圍、統計的單位、逐級報告的秩序、塡報的時間，這些個事項都應該有套硬性的管理規定。

靜態的原始資料的蒐集是以調查法爲主，動態資料則以登記法爲主。登記法和調查法就搜集原始資料而言，是缺一不可的；且登記法有時較調查法常使用到，因爲靜態資料是一特定時期的特定現象的資料，還可以使用調查法來蒐集，而動態資料則是一個現象在一時期內不斷的變動的情況，使用調查法來搜集是不可能的，必須經常作記錄，如此一來，有關單位就必須要不斷的登記所要發表的資料了。

例如學校對於學生出席或曠課或請假的登記，戶政機關對於所轄地區居民遷入、遷出、出生、死亡、結婚、離婚的登記等，都是相當重要且不可缺的，屬於原始資料的動態登記。登記工作的進行，不似調查工作繁瑣，十分方便，因此，登記法對統計資料的蒐集來說，是種事半功倍之法。

第二章
正確的統計資料具有甚大的應用價值

資料的統計是經由調查或登記而獲得的，一份統計資料如果係屬確實可信，是頗具使用價值的。通常我們將統計資料分成許多種類，例如依其來源而分的初級資料和次級資料；依其性質可分成靜態和動態資料；由度量的方法則可分成計數資料以及度量資料，大致就分成這幾類，下述說明之：

①依資料的來源區分：

ⓐ初級資料primary data

直接於資料原來的來源收集而得，所以初級資料也叫作直接資料，這種資料是第一手的資料，尚未經過處理的資料，也就是資料搜集人員最初的調查和登記。

ⓑ次級資料Secondary data

是由他人的單位或機關或者是某個個人曾搜集過，且已經過處理的資料就叫作次級資料，也可稱之爲間接資料。

②依資料性質區分：

ⓐ動態資料Dynamic data

ⓑ靜態資料Static data

動態資料與靜態資料的詮釋已於上章討論過，不再贅述。

③依量度方法分：

ⓐ度量資料Measurement data

如測量道路長度、氣溫、體溫、身高、體重……等利用儀器才能

獲得的資料就叫作度量資料。

ⓑ計數資料Enumeration或Counting data

例如計算一個團體裏男女的比數，或近視與否的比例，這種資料是依照算計對象的特徵或特點或是歸屬某類來區分後，再予以計算，這就是所謂的計數資料。

依照資料的型態也可分為二：

①量性數據quantitative data

也就是資料的性質是可數的，常與變數（variable）或變量（variate）發生關係，例如氣溫、體溫、血壓等的測定，必須用數字來表示的就稱為量性數據。量性數據又可分為：

ⓐ連續型 continuous

連續型數據的單位可以是整數也可以是分數，如肺活量、身高、體重……等都是。

ⓑ間斷型 discrete 或 non-continuous

是間斷型的數據，就必須是完整的一個數，譬如說計算家庭的小孩有多少，每天的營業銷售量是多少，是不能以小數或分數來表示的。

②質性數據qualitative data

不是可數的資料，如一個人的國籍、性別……等都是。

任何資料，如果尚未經過處理，會呈現紛亂無秩序的局面，我們研究者著實不容易在一個紛亂無序的情形下，查出它暗含的某種規律或某一特性。為了達到有系統並簡化的目的，我們就必須加以處理，化雜亂無章為有條不紊的形式，消弭資料數據的繁瑣與雜亂。

通常處理資料有人工整理法和機器整理法二種：

人工整理法：如果資料不是很多，則可採用此法，又可分為卡片整理法和劃記整理法二類，我們將其分述如下：

①卡片整理法Card method

就是使用卡片來歸納資料的方法，有以下幾個步驟：

ⓐ用硬度高一點的紙，製成格式大小相同的紙牌那是卡片。

ⓑ將所搜集得到的資料記載在每張卡片上。

ⓒ檢查歸屬卡片時，有無錯誤發生。

ⓓ然後就是歸類的工作，依每張卡片所記載的性質來歸分。

ⓔ詳檢歸類工作時，有無錯誤發生。

ⓕ計算各類卡片的數量，然後以點計於空白表格上。

卡片整理法的優點是，如歸類有錯誤可很快查出，更正容易，但缺點是既費時又費錢。

②劃計整理法 Tally method , cross - five method

是依據原始的登記與調查表，以人工的方法來歸類，劃記的符號可自定，常用的符號有二種，“正”和“𝍸”。

機器整理法

如果資料的數量繁多，爲了節省時間，應用機器來處理較爲適宜，步驟如下：

①規定號碼：使用機器的特徵就是只用號碼，不用文字，所以，凡是屬於質性資料的，都要改以號碼來替代，將要分類的資料都改變爲數量的型式。例如「1」表示男子，「2」表示女子，近視者以「0」表示，非近視者以「1」表示，文盲以「0」表示，小學程度以「1」表示，中學程度以「2」表示，大專以上以「3」……都是。

②標註號碼：除了數量形式以外的資料，凡調查表或登記表上所記載的內容，都要標註相當的號碼(coding)。

③卡片打孔：把製好的卡片放在打孔機上，按照調查表或登記表內所標註的號碼，在適當的位置打孔，應先行規定各事項在卡片上所佔的部位及先後次序，最好是印在卡片上方。

④卡片分發：把已打孔完成的卡片送入分類機，通過電流的處理，打

孔位置相同的卡片便會歸在一起，而且還能自動計數、複印並歸入早先製好的計數整理表裏，如此，歸類的工作便告完成。

分類和歸類 Classification and Summarizing

資料整理可分三步驟：

①分類：因爲天地之間各種事物都會有所不同，爲了使其相異減至最低，就要分類，經過分類的資料便由繁瑣化爲簡明，因之，分類乃是整理資料的第一步驟。

②歸類：經過分類之後的資料，進行第二步驟，就是把相異處較少或根本完全雷同的併入一類，那麼，繁雜的資料就有系統多了。

③列表：經過分類和歸類後的資料，進行第三步驟，依其特定的規格，次序編列成統計表。如此，原始資料就有條有理了。

依上述三步驟來看，整理一份原始資料，化繁爲簡是最大的目的，整理資料依此三步驟進行，方可達到此效果。通常整理的工作是到列表即告一段落，如有必要，可再行製圖或繪圖，所以整理的步驟，便可以再加上製圖的第四步驟。

進行分類工作的人，將一現象或事物按時間、空間或性質相同的歸爲一類，例如人口，按不同國籍，可分爲中國人、美國人、日本人……。按性別，可分男、女。按膚色，可分黑、白、棕、黃、紅五種。

分類時應注意的二大原則：

①周延：所分類的每一類都要能納入分類，一事物或現象不能納入這一類，就須能納入另一類，所以，不論那一分子，都要能有所歸屬，遺漏一個都不行。

②相斥：也就是所分的各類都不能有模稜兩可的情形，像可歸入甲又可併入乙，各類之間的界限要顯明，不能重複。

這二個原則，理論上談都是相當簡單的，但若實際實行起來，種種的困難就都出現了。

分類時，除了要能把握上述二大原則外，顯示特性也是相當重要的。分類如果過於籠統就失去了分類的價值，如果過於精密細瑣又難完成分類，資料的特性就不易顯示出來。失去了分類的意義。

分類完成後，把原始資料按照所定的分類原則納入所屬各類即是歸類的工作，正確和經濟是歸類的二大原則。歸類時，要使每一原始資料都能正確且毫無遺漏的歸納於所屬各類中，而且要依據原始資料的性質、特性或數量多少，擇一最省時、省力、精確又省錢的歸類法，避免錯誤並達實際效果。

統計表的內容 Contents of A Statistical Table

所謂統計表就是原始資料經過分類與歸類的手續，且依規格所定製成的表格。

無條理、無系統是原始資料的特徵，一旦經過分類歸類並填入表格，便能使人一目了然，可有下列六種功用效果：

①觀察比較便利。

②計算、分析便利。

③記憶便利。

④避使用文字反覆說明的麻煩。

⑤容易查出統計資料內含的特性和規律。

⑥審核方便。

構成統計表的要素：

①標題 Heading

包括表號 Table number、表題 Title、題注 Head note, General note。

②表側 Stub

包括表側首格 Stubhead or box、標目 Center head or subhead、資料行 Line caption or data line。幾個資料行加

上總數行及標目合稱爲段（ Block ）。

③表頭 box head

包括資料欄 Column head or column caption、表頭首格 Spanner head（也就是若干個資料欄的總行、個別的表頭首格和資料欄合稱表頭區（ panel ）。

④表體 Field or body

表側的右方、表頭的下方，是記載統計數字的位置。填列個別數字的位置稱爲細目格 Cell 、橫向稱行 line 、縱向稱欄 column 。

⑤附註 Footnote

表格的下方，爲填寫註解的位置。

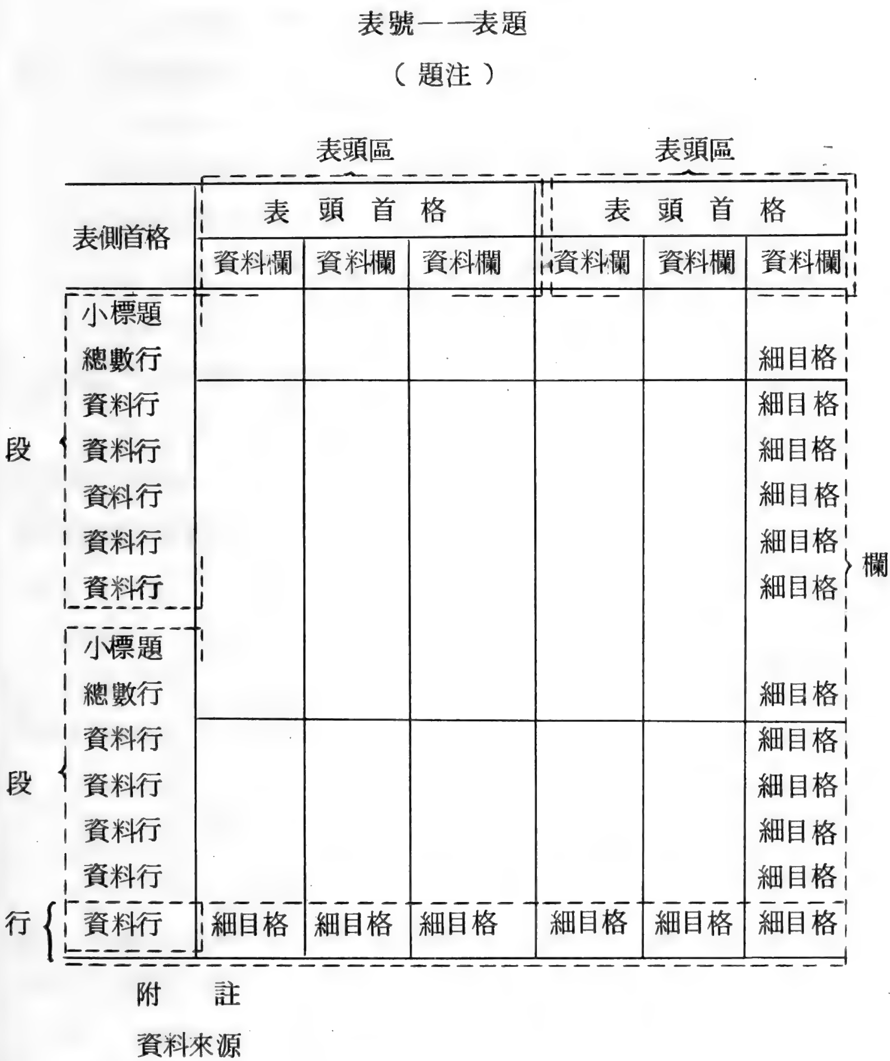
⑥資料來源 Source note

⑦縱線 Vertical 是指分劃表的縱線，橫線 Horizontal rule 是指分劃表的橫線。

依上述的各項名稱列表如下：

- 標　題 Heading
  - 表號
  - 表題
  - 題注
- 表　側 Stub
  - 表側首格
  - 段
    - 小標題
    - 資料行
  - 段
    - 小標題
    - 資料行
- 表　頭 Box head
  - 表頭區 Panel
    - 表頭首格
    - 資料欄
  - 表頭區
    - 表頭首格
    - 資料欄

表體 Field Body
- 行
  - 細目格
  - 細目格
- 欄
  - 細目格
  - 細目格



統計表各部名稱

統計表能否達到眞確，完全視設計的目的與材料的特性決定。表格的製作並非絕對的統一規格，通常有以下幾點原則：

①表題與表體要分開，中間添上橫線也可以，倘若表格較小，各欄的縱線可以省略。

②如果資料是由他人處獲得，附註中必須註明資料來源。

③表格越簡單越好。表若是太大，內容是頗豐富，但過於繁複，不容易一目了然，也就失去了列表的價值。

④表格所表示的爲何，須能自行解釋，爲了此目的要注意：

ⓐ表題要簡單明瞭，包含何事？何時？何地（What ? When ? Where ? ）才是符合正確表題的要求。

ⓑ各行各欄要簡明清楚的表示出。

ⓒ附註中要言明表格中所使用的代號、符號或縮寫各代表什麼。

ⓓ要寫出數據的單位。

ⓔ要列出數據的小計及合計。一般都寫在表格左上方或右下方的位置，端賴表題或組數來決定。如果表有足夠的空白處，總計數可以在左上及右下同時列出。

通常製表的規則如下：

①標題的規則：

統計表的上方要有說明內容的文字，也就是標題，製作標題時應守下列原則：

ⓐ標題的位置是在表的最上方，並且由左至右書寫。

ⓑ標題的用字要簡明扼要，並符合表之內容所述事項。

ⓒ時間、空間與內容的特點，要在標題中顯示出來。

ⓓ如果標題無法以一行盡書，可以小一點的字號寫在標題下，是爲副標題。

ⓔ如果表的內容太多，分成好幾頁，每頁的開頭都要寫出標題，同

時加註「續」或「續前」的字款。

②小標題的規則：

ⓐ小標題要有次序的由上至下、由左至右順序排列。有以下標準：

1. 數量的大小。

2. 地域的位置。

3. 筆劃的多少。

4. 字母的先後。

5. 等級的高低。

6. 重要的程度。

7. 時間的先後。

ⓑ安置標題的位置時，要先權衡輕重，較重要的就安排在表中較醒目的位置，次要的就安排於較不重要的位置。

ⓒ如果一大項目下分若干小項目，或一小項目又分若干小細節，應各間隔一欄書寫。

ⓓ表格中的橫標題，一般是由左至右寫。縱標題則是由上至下排。

③數字的規則：

ⓐ標目下方要註明數字的單位，如公斤、公分等。若表格中的單位完全相同時，可統一寫在右上方即可。

ⓑ同行或同列的數字單位都相同時，不可只以「同上」、「同左」或「〃〃」表示，要個別填上單位名稱。

ⓒ一般表格內的數字大多採用阿拉伯數字。

ⓓ表格內的數字，要上下左右對齊，通常是以個位數或是小數點為基準，若數字多則必須加分節點。

ⓔ若有空格不填任何數字者，可在該位置劃一「—」短線補填，避免查表者懷疑是否漏印或有或無的猜疑。

④線條的規則：

ⓐ表中縱行的縱線都要明顯的一一劃出，橫線就無此必要了，爲了方便查看，也可作每五列空一列的安排。

ⓑ表的邊線要以較粗的線條框出，但若表佔許多頁，則前幾頁的下端可以不必劃邊線。

ⓒ表的左右邊線通常是省略不劃的。

ⓓ總計、平均數、百分數的縱行或橫行，最好以輕細的線劃分。

⑤附註的規則：

ⓐ如果表內的任何部分，如標題、數字單位要加以說明，可在表的最下方，加上附註。須要附註說明的標題、小標題、數字單位，應左右上方標上言明是附註的符號，常以「※」或①②③來表示。

ⓑ表格內的資料來源，取料何處，亦應於附註中註明。

ⓒ附註中的用詞用句要簡明扼要。

第三章
平均數的分類與意義

一、平均數在統計圖上都是用點來表示的

何謂平均數？種類有那些？

將紛亂無序的原始資料，透過整理的手續，使其簡明扼要，就是統計的目的所在。要用簡單易懂的量數來表明全部的資料在表示什麼，方便與其他的資料作比較分析，這也就是平均數之所以產生的原因。

統計學上的集中趨勢（ Central Tendency ）是指，在變量數列中，通常在中間位置的都是變量較多的，所以有此名詞。

平均數就是這些集中趨勢量數的表示，因此又名集中量數（ Measures of central tendency ）。此外，平均數還有點量數之稱，是因為它在統計圖上都是以點示出的。

統計上常用來替代群體的平均數有下列幾項：

①衆數Mode。

②中數Median。

③算術平均數Arithmetic average or mean。

④調和平均數Harmonic mean。

⑤幾何平均數Geometric mean。

任何一種平均數，都分別有它應用的特點和適當應用的情況。

算術平均數

一般就簡稱平均數，是總量數除以總次數所得的商。比如，三個人身高的平均數，就是將此三人的身高相加，總和再除以三，所得之商便是。

①算法

ⓐ從尚未分組的資料求算術平均數

$$M = \frac{\sum_{t=1}^{N} X_t}{N} = \frac{\Sigma X}{N}$$

公式〔3－1〕

M：表示算術平均數的符號

N：總次數

X_t：表示一群量數，$_t$ 是X的附標，即 X_1，X_2，X_3……X_n 各量數，有時也可省略 $_t$ 不用，僅以X表示。

Σ：是希臘字母之一的大寫體，讀音 Sigma，是表示相加的符號，在統計學上是多項數量總和（Sum）之意。

ⓑ由分組資料求算術平均數

1 普通法

$$M = \frac{\sum_{t=1}^{K} f_t X_t}{N} = \frac{\Sigma fX}{N}$$

公式〔3－2〕

X_t 或 X：各組的組中點

f_t 或 f：各組的次數

K：組數

N：總次數，也就是各組的次數和，因此，也可寫成

$$N = f_1 + f_2 + f_3 + \cdots\cdots + f_k = \Sigma f$$

各組資料的總值，並不一定就等於此組中點和次數的乘積，所以兩組的相差是難免的，只要次數分配表的製作無誤，一般說來這個差額是

甚小的，幾乎根本不影響計算的結果。

2 簡捷法

$$M = M' + \frac{\Sigma fd}{N}$$

公式〔 3 － 3 〕

M′：是假設的算術平均數，M′ 可能爲任何數，爲了方便計算，一般是以次數較多，居中間位置的組中點爲M′。

d ：是M′ 與各組組中點的差數，即 $d = X - M'$

公式〔 3 － 3 〕，無論各組的組距相等與否都能應用上。但是，如果各組的組距相等，用下列公式〔 3 － 4 〕則更方便：

$$M = M' + \frac{\Sigma fd'}{N} i$$

公式〔 3 － 4 〕

$d' = (X - M') \div i$，也就是，d′ 是以組距爲單位的各組組中點和假定平均數的差距。

i ：表示組距。

應用公式〔 3 － 4 〕計算算術平均數時，d′ 的數值無須逐一計算，只要在M′ 上寫上零，自零朝左右分別寫1，2，3 ，……及－1，－2，－3，……就可以了。組值如果比M′ 組大的d′ 爲正，小的d′ 爲負，十分方便。

不過，公式〔 3 － 4 〕只限用在組距相等的次數表，雖然組距不相等的次數表可以先化爲組距相等的次數表，再套上公式〔 3 － 4 〕來計算，但此法就多了一道手續，倒不如直接應用公式〔 3 － 3 〕來得方便。

公式〔 3 － 4 〕中的 $\frac{\Sigma fd'}{N} i$ 與公式〔 3 － 4 〕中的 $\frac{\Sigma fd}{N}$ 都是叫作校正數，一般都是用C表示 $\frac{\Sigma fd}{N}$，而以C′ 來表示 $\frac{\Sigma fd'}{N} i$。

$\frac{\Sigma fd'}{N} i$ 計算的步驟，應先計算 $\frac{\Sigma fd'}{N}$ 後再乘以i，或者先計算Σfd′

乘 i 之後再除以 N ？因爲 C′ 的值常是不盡小數的，因此，先計算 $\Sigma fd'$ 乘 i 再除以 N 的答案可能會比先計算 $\frac{\Sigma fd'}{N}$ 再乘以 i 的答案更精確些。

ⓒ加權算術平均數

$$W、M = \frac{\Sigma WX}{\Sigma W}$$

公式〔 3 － 5 〕

W 、M ＝加權平均數

W ：權數

X ：各量數

因爲資料中各個數值的輕重地位不一樣，所以在計算算術平均數時，必須依照它的輕重來加權。

加權平均數是否大於簡單平均數，過是不一定的，要看統計資料的情形爲何來決定。若統計資料的各項數值大小和權數的大小方向一樣，那麼，加權平均數就會大於簡單平均數，若二者的方向不一，加權平均數就會小於簡單平均數，若方向不定，則二平均數一樣。再者，若統計資料項數小，加權計算影響平均數結果大；若項數很多，因爲權數正反相近，所以加權平均數與簡單平均數就會十分近似。

②算術平均數有些什麼特點：

ⓐ算術平均數的求法是由全部的量數求出，是統計上最常用到的平均數，是統計學上的代表。

ⓑ各個量數和算術平均數的差數的代數和爲零，也就是資料中數值較算術平均數爲大的各數和算術平均數相差的和，剛好等於資料中數值較算術平均數爲小的各數與算術平均數相差的和數。如果以符號來表示，可寫成：

$$\Sigma(X - M) = 0$$

公式〔 3 － 6 〕

也可寫成：

$$\Sigma(X_t - M) = \Sigma(M - X_s)$$

公式〔3－7〕

X_t：數值大於M的量數

X_s：數值小於M的量數

ⓒ只要知道總數量和總次數，就可以算出算術平均數。

ⓓ算術平均數符合代數運算的規則，若算術平均數有二個以上，也可合併求出總平均數。

$$M = \frac{N_1M_1 + N_2M_2 + N_3M_3 + \cdots\cdots}{N_1 + N_2 + N_3}$$

公式〔3－8〕

ⓔ算術平均數因受抽樣的影響很小，所以在統計資料中可說是最準確的一種。

ⓕ算術平均數受極端量數的影響較大，若有此類情況，最好是放棄採用此法，可改用中數或衆數代替。

中位數

全部的量數，依其大小順序排列後，居於中間位置的就是中位數，也可說量數標尺最中間的一點就是中位數，以此點爲中點，上下各有二分之一的量數。根據中位數的所處地位，因之又稱地位平均數。

①算法

ⓐ由未分組資料中求中位數

M_e 的位置$= \dfrac{N+1}{2}$

M_e：中位數

N：總次數

ⓑ由分組資料求中位數，各組量數平均安排在組距中，這和求M視各組量數密集組中點的不一樣，公式如下：

$$M_e = L + \frac{\frac{N}{2} - F}{f} i$$

公式〔3－9〕

或：

$$M_e = U - \frac{\frac{N}{2} - F'}{f} i$$

公式〔3－10〕

L：中位數組的低限

U：中位數組的高限

f：中位數組的次數

F：組值小於中位數組之各組的次數和

F'：組值大於中位數組之各組的次數和

公式〔3－9〕是由數值較小的一邊開始計算中位數，公式〔3－10〕則是由數值大的一方開始計算中位數。

公式〔3－9〕和公式〔3－10〕是凱萊（Kelley）根據前學者范契奈（Fechner）的公式演變而來，所以也叫作凱萊中數公式。

我國也有一耿相曾氏依凱萊公式又演化出如下公式：

$$M_e = M_e' + \frac{F' - F}{f} \times \frac{i}{2}$$

公式〔3－11〕

M_e'：是 M_e 所在組的中點，其他的符號所代表的都和上述相同。

上述三個公式〔3－9〕〔3－10〕及〔3－11〕中，以耿相曾氏所變換而得的公式較好，因爲初入門統計學的學人，常受計算算術平均數簡捷法的左右，把凱萊公式中的L或U寫成中位數組的組距中點，而發生錯誤結果，若使用公式〔3－11〕來計算中位數，和計算算術平均數簡捷法公式相若，可避免發生錯誤。

②中位數的特點

ⓐ組距或組限的變化對中數的影響較少，所以位置較穩定。

ⓑ中數因不受極端量數存在的影響，所以數值較固定。

ⓒ中數與各量數之差數絕對值的總和最小。這個特點可用下式表示：

$$\Sigma \ |X - M_e| < \Sigma \ |X - X'|$$

公式〔3－12〕

ⓓ中位數並不是從全部的量數中求得，因此，它的正確性就不如算術平均數，且比算術平均數不穩定，也不符合代數的方法求取結果。

ⓔ只知道次數而不知量數的值時，還是可以求出中位數。分配的一端或二端的組距沒有特定的範圍限制，中位數是最可靠的平均數。

③其他分割數

中位數是把統計數列按大小順序排列後，從中點將全部數列分成二部分，大於中位數與小於中位數的量相等，所以我們說中位數也就是分割數。

分割數除中位數之外，還有四分位數、十分位數和百分位數。四分位數就是將中位數二端的數量再次的分割爲二，由數值小的一端算起，我們分別稱它爲第一四分位數（ First Quartile ），又叫作下四分位數（ Lower Quartile ），後半部的分割點稱第三四分位數（ Third Quartile ），又叫作上四分位數（ Upper Quartile ），所以中位數也就是第二四分位數。中位數既是將數列分成二部分，四分位數是將數列分成四部分，依此類推十等份的分割就是十分位數，百等份的分割就是百分位數……。中位數化爲四分位數時是第二四分位數，中位數化爲十分位數時，是第五個十分位數；中位數化爲百分位數時，是第五十個百分位數。Q_1 是表示第一四分位數的符號，Q_3 是第三四分位數，D_1 是第一個十分位數，D_2 是第二個十分位數，D_k 是第 k 個十分位數，P_{10} 爲第十個百分位數，P_{15} 是第十五個百分位數，P_p 是第 p 個百分位數。

四分位數、十分位數、百分位數的計算法和中位數的相同。沒有分組的數列，是把數值由小至大依順序排列後，先求出需要的分割數的位置，再求出將位數的數值就可以了。

Q_1 的位置是（$\frac{N}{4}+\frac{1}{2}$）之處，Q_3 的位置是（$\frac{3N}{4}+\frac{1}{2}$）之處，D_R 的位置是（$\frac{RN}{10}+\frac{1}{2}$）之外，$P_p$ 的位置是（$\frac{RN}{100}+\frac{1}{2}$）之處，就跟計算中位數先求出它位置在（$\frac{N}{2}+\frac{1}{2}$）處相同。若資料已歸類分組完成，就可應用中位數的公式，換算爲四分位數、十分位數或百分位數的公式：

$$Q = L_1 + \frac{\frac{N}{4} - F_1}{f_1} i$$

公式〔 3 － 13 〕

$$Q_3 = L_3 + \frac{\frac{3N}{4} - F_3}{f_3} i$$

公式〔 3 － 14 〕

$$D_R = L_R + \frac{\frac{RN}{10} - F_R}{f_R} i$$

公式〔 3 － 15 〕

$$P_p = L_p + \frac{\frac{PN}{100} - F_p}{f_p} i$$

公式〔 3 － 16 〕

④百分等級

由上述計算百分位數的公式，便知百分位數可還原爲原來數值，同樣的，也可自原來數值來計算它的百分位數。

超過全部資料的百分比數就是百分位數。因爲它明白的顯示原始數值在百分量尺上的等級或地位，所以這個百分數又叫作百分等級。

ⓐ由數量資料求百分等級

$$P_R = \frac{100}{Ni}\left[f (x - L) + Fi \right]$$

公式〔3－17〕

P_R ：代表百分等級

N ：總次數

i ：組距

x ：是計算百分等級的數量

L ：含X的一組之低限

f ：含x的組一的次數

F ：小於x組的各組之次數和

ⓑ由品質資料求百分等級

公式如下：

$$P_R = 100 - \frac{100R - 50}{N}$$

公式〔3－18〕

R：需計算百分等級的其中一等級

N：總等級數

二、調和平均數爲各量數倒數之算術平均數的倒數

調和平均數

各量數倒數的算數平均數的倒數，即是所謂的調和平均數，所以也叫作倒數平均數。

①算法

$$H = \frac{1}{\frac{1}{N}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots\cdots + \frac{1}{x_n}\right)} = \frac{N}{\Sigma\left(\frac{1}{x}\right)}$$

公式〔3－19〕

H：調和平均數

x：各量數

以幾何平均數來度量物價的波動情況是最精確的方法。

②調和平均數使用價值

調和平均數若應用到生物統計和人口統計時，價值不高，但若在其他統計任何一特定情況之下，能有計算平均速率、平均滙價與平均物價的功能。

求平均速率，應用調和平均數或算術平均數，必須視不同的情況來決定。

計算時間速率有二個要素，一是以 t 表示的時間、另一是以 d 表示的工作量或距離。速率也有二個代表方式，一爲 $\frac{d}{t}$ 的方式，另一則爲 $\frac{t}{d}$ 的方式。在什麼情況下求調和平均數，什麼情況下求算術平均數，有下述原則供參考：

① 如以 d 代表固定元素（距離或工作量相等），速率以 $\frac{d}{t}$ 方式表示的情形下，或者以 t 代表固定元素（時間一樣），速率用 $\frac{t}{d}$ 表示的情形下，皆用調和平均數計算平均速率較適宜。

② 如果以 d 代表固定元素，速率用 $\frac{t}{d}$ 表示的情況下，或以 t 代表固定元素，而 $\frac{d}{t}$ 方式表示速率之情況下，最好是利用算術平均數計算平均速率。

計算物價有兩要素，一是物量，以 q 表示，另一是貨幣的數量，以 v 表示。表示物價有二個方法，一是 $\frac{v}{q}$，設 v 爲固定元素（貨幣數一樣），物價用 $\frac{v}{q}$ 表示，或者是以 q 爲固定元素（物量相等），物價用 $\frac{q}{v}$ 表示，使用調和平均數計算平均物價較合宜。若以 v 爲固定元素，以 $\frac{q}{v}$ 表示物價，或者以 q 爲固定元素，以 $\frac{v}{q}$ 表示物價，則應用算術平均數來計算平均物價較好。

計算滙價的要素有二，一是滙費，用 m 表示；另一是以 r 表示的滙款。表示滙價也有兩個方法，一是 $\frac{r}{m}$，另一爲 $\frac{m}{r}$；假設固定元素用 r 來表示，滙價用 $\frac{r}{m}$ 來表示的情形下，或者是固定元素用 m 來表示，滙價用

$\frac{m}{r}$來表示的情形下，利用調和平均數來計算滙價較適宜。若固定元素用 r 表示，用$\frac{m}{r}$來表示滙價的情形下，或者是固定元素用 m 表示，用$\frac{r}{m}$來表示滙價，則應用算術平均數來計算平均滙價才適宜。

當計算平均速率，t 及 d 都不是固定元素，計算平均物價時 v 和 q 都不是固定元素，計算平均滙價時，m 和 r 都不是固定元素，就要個別求出總數後再相除。

③調和平均數的特性

ⓐ若數量中有零或負數的情況下，就無法求出調和平均數。

ⓑ二個正數的調和平均數和算術平均數相乘所得的積數，與其幾何平均數的乘方相等，以符號表示：

$$H \cdot M = G^2 \quad or \quad G = \sqrt{H \cdot M}$$

ⓒ算術平均數大於幾何平均數，幾何平均數又大於調和平均數，以符號表示：

$$M > G > H$$

（若數列中全部數值一樣，則 $H = G = M$ ）

ⓓ調和平均數可以代數方法計算，若知各部分數列的項數與調和平均數，便能算出全部的調和平均數。

$$H = \frac{N_1 + N_2 + \cdots\cdots + N_k}{\frac{N_1}{H_1} + \frac{N_2}{H_2} + \cdots\cdots + \frac{N_k}{H_k}}$$

公式〔 3 － 20 〕

ⓔ若數量已經歸類，或者須加權的情況下，也可計算它的調和平均數。

$$H = \frac{f_1 + f_2 + \cdots\cdots + f_k}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots\cdots + \frac{f_k}{x_k}} = \frac{N}{\Sigma\left(\frac{f}{x}\right)}$$

公式〔 3 － 21 〕

$$W \cdot H = \frac{W_1 + W_2 + \cdots\cdots + W_n}{\frac{W_1}{x_1} + \frac{W_2}{x_2} + \cdots\cdots + \frac{W_n}{x_n}} = \frac{\Sigma W}{\Sigma\left(\frac{W}{x}\right)}$$

公式〔 3 － 22 〕

各平均數的使用

①每種平均數的準確程度

ⓐ計算算術平均數，要先算出各量數的總和（ΣX）。再除以總次數（N）。因爲N大多是正確數，所以，算術平均數要和ΣX的數值的有效數字位數一樣。

如果以簡捷計算算術平均數，無法確知各量數的總和的有效數字，此時，必須先計算出算術平均數，以便使它的正確度和原來量數中的有效數字位數最少的相等，再乘以N，算出的有效數字就是算術平均數應當有的有效數字。

ⓑ衆數和中位數都是自點數（ Connting ）中獲之，粗看之下似乎是精確值，但是因爲受到歸類分組的關係，一般都用近似值來處理，一般衆數之有效數字，大多不會超過原來數量的有效數字，中數的有效數字大多不會大於算術平均數的有效數字。

ⓒ調和平均數的計算，是依照一數列的倒數之和來計算，縱使一個數值的倒數的有效數字跟原來數值的有效數字相等，倒數總和的有效數字也可一定和原來數值總和的有效數字一樣，基於方便的原則，一般調和平均數的有效數字，大多不會大於算術平均數的有效數字。

ⓓ計算幾何平均數是依照一數列的連乘之積來計算，它的有效數字，按理不會大過原來變量最小的有效數字。

②平均數的使用選擇

每種平均數都有它的缺點和優點，要選擇何種使用必須詳加研究，依G. U. Yule提供六種選擇優良平均數的要點：

ⓐ內含簡單明確。

ⓑ同一資料計算同一平均數，要能有相同的結果，也就是要確定確實。

ⓒ能夠符合代數法則的整理，還原後能夠相符，或者從若干平均數與項數，都能計算全部的平均數。

ⓓ一數列中任何一環有變異時，平均數也要有變動。

ⓔ容易計算。

ⓕ儘量不受抽樣改變的影響，由同一資料中選擇若干樣本，算出的平均數差數越少越好，也就是平均數越穩定越好。

算術平均數完全符合上述六項要點，而且只要知道一數列的項數與總和就能算出算術平均數。但是某些時候算術平均數受到極端數值影響太大就不宜使用了。此外，分組資料中組距不確定時，也不能使用算術平均數，應改用他種平均數。

一般說來，衆數與中位數沒有算術平均數來得實用，因爲不適合代數方法的運用，並且靈敏度也不夠。但是它們的優點是可以不受到兩極端數值的影響，當次數分配表中的組距不確定時，能很快的求得，並且適合質與量的資料。

調和平均數與幾何平均數，因爲計算麻煩，不容易讓人理解，除非是特殊情形需要，否則通常很少用到，所以是一種補助的平均數。

無論那種平均數，都有它適用的情況，應用時，端賴資料性質選擇最合適的平均數。

三、衆數係一個數列中出現次數最多的數值

衆數

一個數列中出現次數最多的就是衆數，衆數在次數分配表是次數最多的一組的中點，又於次數分配圖上在橫座標上，最高點的數值也就是

衆數。這種衆數往往隨著組距與組限的變異而不同，欠缺安穩性，稱爲觀察衆數或大概衆數。

在統計學的理論上，理論衆數乃是理論曲線（也就是常態曲線）縱線上最高點的數值，因爲理論衆數的計算十分複雜，一般都採用衆數。因觀察衆數缺乏正確性，可應用下述方法計算出近似衆數：

①算法

ⓐ皮爾生（K. Pearson）之經驗法

$$M_o = M - 3(M - M_e)$$

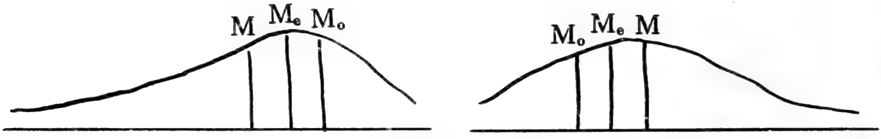
公式〔3－23〕

M_o：代表衆數

M　：代表算術平均數

M_e：中位數

英國學者皮爾生根據經驗獲知，單峯微偏的分配圖中，衆數、算術平均數與中位數三者常保持一種固定的關係，中位數一定位於算術平均數和衆數之間，此外，算術平均數和衆數的距離往往是中位數距離的三倍，因此發明上述公式。



M，M_e，M_o 在單峯微偏分配圖中的位置關係

ⓑ金氏（W. I. King）的插補法

$$M_o = L + \frac{f_2}{f_1 + f_2} i$$

公式〔3－24〕

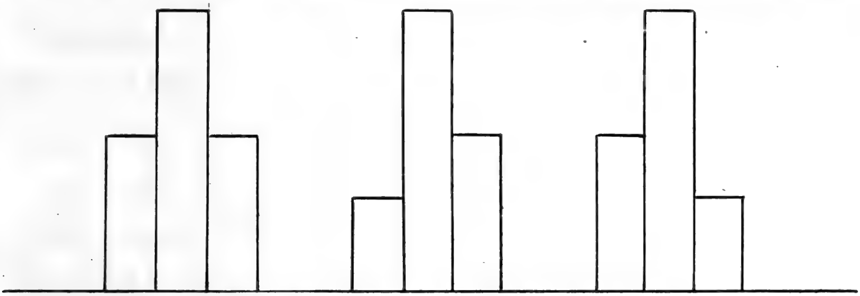
L：衆數組之下限

f_1：組值小於衆數組之相鄰組的次數

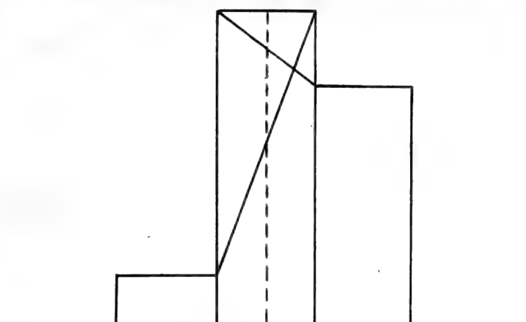
f_2：組值大於衆數組之相鄰組的次數

其餘符號與上述同。

衆數在衆數組內的位置與上下二組限的距離，決定於它相鄰兩組的次數之反比，相鄰兩組的次數相同時，衆數在衆數組的位置距離它的上下限是一樣的話，那麼，衆數組的組中點就是衆數。如果相鄰二組的次數不相等，若組值較大組次數較多的情況下，衆數在衆數組的位置距離高限較近，也就是說衆數的數值較衆數組的中點要大。相反的，若組值較小組的次數較多時，衆數在衆數組的位置距離下限較近，也就是說，衆數的數值較衆數的中點小。上述的公式就是根據這些原則而創立的。



衆數位置及其相鄰兩組次數的關係圖



衆數位置及衆數組次數與相鄰兩組次數的關係圖

ⓒ克魯伯氏（E. Czuber）的比例法

$$M_o = L + \frac{f_1 - f}{f_1 + f_2 - 2f}\, i$$

公式〔3－25〕

f：衆數組的次數

其餘符號同上。

公式〔3－24〕決定衆數的位置是以衆數組二相鄰的位置，克魯伯氏認爲衆數組本身的次數之於衆數的位置也是相當重要的，上述公式因而創立。

應用金氏法與克魯伯氏法所求得衆數相差很小，但皮爾生氏法求得的衆數與金氏法及克魯伯氏法的差異頗多，這是因爲皮爾生氏法是由經驗獲得，除了單峯微偏時計算衆數較合適以外，無法普遍應用。計算衆數若不應用觀察衆數，金氏法是輕易行的。

ⓓ併組法

在次數分配表中，如果有一組次數特別多，使用皮爾生法、觀察法、金氏法、克魯伯氏法求衆數都很容易。不過，如果次數分配表中沒有明顯的衆數組時，應用併組法計算衆數較爲利便。

二組合併時，先將一二兩組加起來，三四兩組加起來，五六兩組加起來，然後相加二三組，四五組……。

三組合併時，首先將一二三組相加，四五六組相加……，其次相加二三四，五六七……，第三次相加三四五、六七八的次數。若三組合併後，仍不能決定衆數的位置，就要做四組以上的合併，一直到能夠確定衆數的位置爲止。

除此之外，應特別注意的，若一次數分配表中有二個衆數的話就是雙峯分配，二個以上就是多峯分配，此時，不易有明顯的單獨衆數可代表。雙峯或多峯分配的產生多是因爲資料過少，受到抽樣變動的影響，

或是資料本身缺乏相同性質所致，若屬前者，修勻法可以減輕它的不規則性並還原到原來的特性，若屬後者，就必須在說明時加以詮釋，才能了解。

②衆數的性質

ⓐ因爲衆數不受極端量數的影響，所以適合用衆數爲概略均數，當兩極端量數相差很大時。

ⓑ數列中出現次數最多的數值就是衆數，計算容易，只要有一部分密集資料就可算出，常爲人廣泛應用。在單峯對稱分配中，衆數、算術平均數、中位數此三者是合而爲一的。

ⓒ如果資料的次數少，衆數的價值就低了。

ⓓ衆數容易受到抽樣變動及組距和組限變動的影響，穩定性不夠，而且也不適合代數的計算，少用爲宜。

幾何平均數

幾何平均數就是若干個數值連乘積的若干次方根，算術平均數、中位數與衆數三者雖是常使用的三種平均數，但某些情況，如計算複利率、某一特定時間的人口數、計算學習者的平均進步率，就不合用了，須改用幾何平均數。

①算法

$$G = \sqrt[n]{x_1 x_2 \cdots\cdots x_n}$$

公式〔 3 － 26 〕

G：爲幾何平均數

x_1，x_2，……：表示各個量數

應用公式〔 3 － 26 〕時，一般以對數計算，或可寫成下式：

$$\log\ G = \frac{1}{N}(\log\ x_1 + \log\ x_2 + \cdots\cdots + \log\ x_n)$$

$$\log\ G = \frac{1}{N}\Sigma\ \log\ x$$

公式〔 3 － 27 〕

由公式〔 3 － 26 〕可知幾何平均數的對數，就是各量數的對數的算術平均數，所以幾何平均數也叫做對數平均數。求出幾何平均數的對數後，再求它的反對數便求出了幾何平均數。

統計資料中，各數值的重要性不一時，也可以求它的加權平均數，公式如下：

$$W \cdot G = \Sigma W \sqrt{x_1^{W_1}\ x_2^{W_2} \cdots\cdots x_n^{W_n}}$$

公式〔 3 － 28 〕

$$\log\ W \cdot G = \frac{W_1 \log x_1 + W_2 \log x_2 \cdots\cdots + W_n \log x_n}{W_1 + W_2 + \cdots\cdots + W_n}$$

$W \cdot G$：表示加權幾何平均數

W_1，$W_2 \cdots\cdots W_n$：代表各數值的權數

$\Sigma W = W_1 + W_2 + \cdots\cdots + W_n$

計算幾何平均數時，若統計資料已分組，各組的組中點就是 x，各組的次數也就等於權數，公式如下：

$$G = \sqrt{x_1^{f_1} x_2^{f_2} x_k^{f_k}}$$

公式〔 3 － 29 〕

$$N = f_1 + f_2 + f_k = \Sigma f$$

$$\log\ G = \frac{1}{N} (f_1 \log x_1 + f_2 \log x_2 \cdots\cdots + f_k \log x_k$$

$$= \frac{1}{N} \Sigma (f \log x)$$

公式〔 3 － 30 〕

②幾何平均數的作用

ⓐ計算比例或百分比的平均

1.計算物價指數

2.計算平均進步率：學習進步的比例，是相同比率的幾何級數，所以在計算平均進步率時，也應同時計算幾何平均數才對。

ⓑ計算量數的平均率

$$x_{\frac{n}{2}}=\sqrt{x_o x_n}$$

公式〔 3 － 31 〕

x_o ：代表前期的量數

x_n ：代表後期的量數

$x_{\frac{n}{2}}$：代表中間時期數量

求複利率與複利本息，也應使用幾何平均數。

©幾何平均數的特點

1.數量中若有一數爲零，便無法用幾何平均數求出。

2.數量中若有一數爲負數，幾何平均數便無意義。

3.適合代數計算，各部分的項數與幾何平均數已知，便可計算出全部的幾何平均數，公式如下：

$$G=\sqrt[N]{G_1^{N}\,G_2^{N2}\cdots\cdots G_k^{NK}}$$

$$N=N_1+N_2\cdots\cdots+N_k$$

公式〔 3 － 32 〕

4.一個數列的幾何平均數永遠不大於算術平均數，可用符號$M_o>G$表示（當各數值全部相等時，$M=G$）。

5.兩數列項數相同時，對應值之比的幾何平均數與兩數列之幾何平均數之比相等，如用符號表示，公式如下：

$$G=\sqrt[N]{\frac{x_1}{x_1'}\ \frac{x_2}{x_2'}\cdots\frac{x_N}{x_N'}}=\frac{\sqrt[N]{x_1x_2\cdots x_N}}{\sqrt[N]{x_1'x_2'\cdots x_N'}}$$

公式〔 3 － 33 〕

第四章
統計圖係一種平面或立體的圖形

一、作統計圖必須遵守的原則

統計圖可以是平面的，也可以是立體的，是應用點與線及條的長短、粗細、多寡、或是所佔面積的比例以及色澤的不同、線條的安排來顯示統計表中各數字間的相互關係。

統計表的功用雖可化繁複爲簡明，但由於表中全以數字代表，必須要仔細查看，謹慎研讀方能有所概念。統計圖就方便多了，因爲它是應用點、線、面、體來顯示變異的情形，使人們一看就能了悟，所以統計圖廣爲各種工作研究者的普遍利用。

製圖應守原則：

①統計數列的種類繁多，故選擇製圖形式應視其性質需要。

②由於統計圖易於使人一目了然，故製圖必須正確可信，否則易使人誤解。

③統計圖製作時應確守簡明之原則，否則就失去了它的製作目的。

統計圖的內容構成要素包括如下：

①標題Title

也就是指本圖的名稱，標題要含括時間（ When ）、地點（ Where ）、及敍述何事（ What ）三重點。

②頭注General　note

標題之下要載明查看圖表的應注意事項。

③比例尺 Scale

比例多少（ Scale numeral ）要記載在圖表旁，說明（ Scale caption ）就直述於比例數下方。

④圖例 legend or key

是指詮釋圖的作法，可載於圖中空白處。

⑤圖野 Field bodv

就是圖中顯示資料的部分，是圖中主體。

⑥橫、縱線 Grid

顯示圖內容的線條，直的是縱線 Vertical ruling、橫的就是橫線 Horizontal ruling。

⑦資料來源 Source

資料的搜集取自何處，要在附註中說明。

⑧附註 Footnote

常安排在圖的最下方，記載注意事項之用。

⑨圖框 Frame

顯示圖的範圍。

統計圖的種類

①條狀 Bar diagram

是統計圖中最簡單的型式，通常用來比較質性資料（ qualitative data ）或是分散的量性資料。

條狀圖包括垂直及水平兩種圖型，以不同的形狀表示不同的數據。特性及應注意事項如下：

ⓐ有關時間者，通常是自左至右順序排列。

ⓑ有關尺度的資料圖不適宜採用此條狀圖。

ⓒ性質不同的直條，間隔要大些。

ⓓ各線條的寬度要一致。

ⓔ基準點應自零（ 0 ）開始。

ⓕ各線條的寬度要相等，且比條與條之間的間隔寬。

ⓖ基準線的條體要粗黑些。

ⓗ第一和最後之間的間隔大抵爲其他中間間隔的一半。

ⓘ圖中若要表示兩種以上不同性質的說明，同時間或同情事的線條要儘量靠近，或者局部重疊。

ⓙ解說線條的說明要寫在圖的下方。

ⓚ圖中若有二種以上的分類，尺度數可分別註於左右二側。

ⓛ線條的安排順序通常是根據其數之大小繪製。

ⓜ圖中若需加入外文，應縱向自上而下書寫。數字就可雙向書寫。

ⓝ水平的條形圖、尺度與說明可書於圖表上方。

②扇狀圖 Pie diagram

以弓形所佔位置來比較數值的大小。個別數值要先算出全部數值中所佔的比例，再乘上 360°，將個別數據以圓形圖來標示。

扇狀圖的特點與應注意事項：

ⓐ圖中文字的字號要一致，並順一定方向填寫。

ⓑ製圖時，應自中間正方上的一點開始。

ⓒ依大小順序順時鐘方向排列。

ⓓ可加註文字輔助說明，但切忌繁複冗長，以免失去圖表的意義。

③百分率條形圖 Proportional bar diagram

若欲顯示的種類不多可用百分率條形圖來表示。此圖中一個長方形代表 100％，依各分類分割成若干段（ sections ）。

④線形圖（ Line diagram ）

以線條來表示變數的變化，主要在於高度的變化，不涉及圖面積。

線形圖的特點及注意事項：

ⓐ縱座標與橫座標均明顯標出，但圖中的線條必須比座標線更顯明，方便查閱。

ⓑ尺度的數值一般都是記在左方或下方，如有必要也可在右方及上方標示。

ⓒ繪製圖基準點應由零（ 0 ）開始，線條的延展也是由左向右延伸。

ⓓ圖中若須用到兩種以上的線條，必須使用不同的曲線加以區別。如實線、虛線或不同顏色的曲線。

ⓔ可於左右兩側標示不同的尺度。

⑤次數多邊圖 Frequency polygons

若欲比較二個以上的次數，可使用次數多邊圖，以彌補直方圖形常相互重叠的缺失，觀察數值假定在各組距的中點，在此位置標一記號，最後再以直線連接就完成了。

⑥柱狀圖（ Column diagram ）

表示連續量性數據，可用此類圖表來顯示。以互相銜接的長方形所構成的面積圖來表示個別數據的大小。各組頻數面積之和要與全部頻數的構成面積相等。頻數的分佈若以百分比或相對頻數代替，組距一樣，製圖時便可直接繪出。若組距不相等，要先變換使其相等再製圖，否則易生弊端。

直方圖的特點及應注意事項

①製圖須能使人易懂，合乎事實。

②軸的基準點應自零（ 0 ）算起，記錄次數的變化。

③橫軸記錄組距的變化。

④附註除載明資料來源外，最好也能說明製圖所依據的統計表。

⑤記上相對頻數，方便比較二個以上的同形圖表。

⑥合宜決定兩軸的比例，以免使製出之圖顯然不襯。（ 過長或過短，過高或過低 ）

二、比較兩組以上的數字應著重變數的相對變化

相關圖Scatter，correlation diagrams

用來表示兩個變數之間的關係，以兩個軸各代表不同的變數。如果兩變數之間的關係是呈一直線，卽表示二者有直線性質之關係，若不成一直線則表示二變數之間的關係不單純，若圖中的標示點成分散形狀，則這些變數可能無任何同性質之關係。

其他圖形

資料圖形的種類繁多，除了上述各圖形外，還有許多專爲特定需要而製作的圖表，依情形需要而定。例如有：

①顯示某地區的地形、土質、或農作物分佈情形的地圖。

②顯示特殊事件發生的地區，以圖表示的針點圖形（Pin map）。

③較易引人注目的像形圖（Pictogram），應用要表現的種類的形象繪製。

④人口構成圖（Population pyramid）。

⑤立體圖表 Stereogram。

⑥計算圖表 Nomogram。

統計圖的選擇必須適合表達的目的，不同的圖形代表不同的意義，也有它適用的情況，並非每一圖形都適用任一記事，也並非任一資料都能應用任一圖形來表達，爲了達到使用便利的目的，圖形的選擇應謹愼處理。

圖表種類繁多，但所顯示的內容，不外下列數項：

①內函的比較，例如所有人類中，金髮、黑髮、紅髮的比率各是多少？

②一部份在統體上所佔範圍的多少？

③時間的變化與事件的關係，如去年死於車禍的人數與今年的比較。

④分佈情形，如一國的農產品各佔全農作物的多寡。

⑤地區的分佈，如蘋果產量最多在那一國，次多是那一國……。

⑥地點分佈，例如信奉佛教的人數最多的是那一國，次多的又是那一國。

⑦如高血壓患者的平均年齡與性別上有無關連的相關問題比較。

第五章
敍述統計及統計推論

一、變異數推論法則

變異數的推論

設Q爲一常態分佈的變異數，從中抽取樣本n，該變異數暫設爲S^2，証驗$Q^2 = Q^2$。可利用公式：

$$x^2 = \frac{(n-1)S^2}{Q_0^2}$$

也可寫成：

$$\frac{\Sigma(x-\overline{x})^2}{Q_0^2}$$

公式〔5－1〕

再以計算獲知之數值比對$x^2(n-1)$分佈卽可。

若採用雙邊檢定法，所得x^2數值應比對表列$1-\frac{1}{2}\alpha$及$\frac{1}{2}\alpha$的機率。

兩平均値之比較

此所分析的是配對問題與非配對問題。配對問題是指兩組樣本的數量相同，並且各分子都能相互配對。非配對問題就是兩組完全不相干，各自獨立。

配對資料（Paired Case）

設有兩組樣本，大小同爲n，一爲：

$$x_{11}\ ,\ x_{12}\ ,\ \cdots\cdots x_{1i}\ ,\ \cdots\cdots x_{1n}$$

是從一常態分佈中抽樣的，平均值是 μ_1 ，變異數是 Q_1^2 。

另一為：

$$x_{21}\ ,\ x_{22}\ ,\ \cdots\cdots x_{2i}\ ,\ \cdots\cdots x_{2n}$$

同是由一常態分佈中抽樣，平均值為 μ_2 ，變異數為 Q_2^2 。

假設上二樣本中的 x_{1i} 與 x_{2i} 是與某事物相互配對，一般情況下，這二組中的每一分子都會有相同的一對配對，也就是數大者與數大配對，數值小的就和數值小的配對（ x_{11} 與 x_{21} 相配， x_{12} 與 x_{22} 相配），若有相同兩數與一數配對，則為：

$$E\ (\ x_{1i}-x_{2i}\ )=\mu_1-\mu_2$$

但 var（ $x_{1i}-x_{2i}$ ）便小於 $Q_1^2+Q_2^2$ ，兩平均值之差應為 $\bar{x}_1-\bar{x}_2$ ，也就是表示若干個 $x_{1i}-x_{2i}$ 的平均差異值，各差異值無關。

若有二組樣本的分子數相同，但二者之間並無可配對者，仍可比較其平均值，只要依照相稱的分子加以整理安置配對，然後再使用七檢定法分解各組的相差值即可，但須注意：

①由於配對排列組合的方法很多，而各種方法排列出的結果並不一定完全相同，甚或會有很大差異之處。

②若樣本的分子個數不同，即不能使用此法處理。

非配對資料 Unpaired Case

例如自一分佈資料中抽取樣本 x_{11} ， x_{12} ，⋯⋯ x_{1n1} ， μ_1 為平均值， Q_1^2 為變異數。另一組中抽取樣本 x_{21} ， x_{22} ，⋯⋯ x_{2n2} ， μ_2 為平均值， Q_2^2 為變畢數， $\bar{x}_1$ ， $\bar{x}_2$ 代表樣本平均，若此二組樣本的分子彼此都不能配對，依抽樣誤差的理論知：

$$E\ (\ \bar{x}_1-\bar{x}_2\ )=\mu_1-\mu_2$$

$$\text{var}\ (\ \bar{x}_1-\bar{x}_2\ )=\frac{Q_1^2}{n_1}+\frac{Q_2^2}{n_2}$$

若 x 的分佈是屬常態，並且 Q_1^2 與 Q_2^2 都是已知數，上式便可推算 $\mu_1 - \mu_2$ 的變化。若要檢定虛無假說 H_0 ：$\mu_1 = \mu_2$ ，標準化常態偏量的計算則爲：

$$Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{Q_1^2}{n_1} + \frac{Q_2^2}{n_2}}}$$

並可計算 $\mu_1 - \mu_2$ 的可信範圍。

若樣本的數量很多，x 分佈的常態性就會受到相當的局限。因爲 $\overline{x}_1 - \overline{x}_2$ 的分佈與 $\overline{x}_1$ 及 $\overline{x}_2$ 的分佈皆近於常態分佈。唯一的問題是 Q_1^2 和 Q_2^2 的大小是多少並不知道，所以須要採用一些方法去估計變數，這個問題可分爲二種情形來研究，一是設 Q_1^2 和 Q_2^2 是相同的二數，並假設這二數有一個共同的估計值。二是完全沒有第一種情況的條件。

①變異數相同：雙樣本七檢定法

假設的 $Q_1^2 = Q_2^2$ 是可成立的，檢定二樣本乃是具有相同平均值和變異數的分佈。假設平均值相同，變異數也相同，檢定這個虛無假說要視 $\overline{x}_1 - \overline{x}_2$ 的差異性質才能分析對 x 平均值的影響。

某些情況下，根據一般的經驗便知如何區分二組樣本，差異變化可能對平均值有影響，但對變異數就沒有顯著的影響，如樣本變異數估計值 S_1^2 與 S_2^2 相差或許很大，但一般情況下，其差異可視爲是因 S_1^2 及 S_2^2 的抽樣有偏誤所造成的，並非 Q_1^2 及 Q_2^2 二者眞有差異。

如果假設 Q_1^2 與 Q_2^2 相同，其共同值用 Q^2 來表示，研究如何估計 Q^2 。

在第一樣本得

$$S_1^2 = \frac{\Sigma_1 (x - \overline{x}_1)^2}{n_1 - 1},$$

Σ 右下方的 $_1$ 代表第一樣本。

第二樣本得

$$S_2^2 = \frac{\Sigma_2 (x - \bar{x}_2)^2}{n_2 - 1}$$

可將 S_1^2 與 S_2^2 二數平均便可算出上述二樣本的共同變異數，但求加權平均才是正規的計算法，如此可減少樣本數不一所造成的差異，所以：

$$S^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$= \frac{\Sigma_1 (x - \bar{x}_1)^2 + \Sigma_2 (x - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

這個手續在計算上是很實用的，因為如果使用 t 檢定法就要考慮到自由度的問題，這裏自由度是 $n_1 + n_2 - 2$，如果 x 是屬常態分佈，上式是正確的解法，但如果 x 的分佈是非常態的分佈，同樣是一近似的解法。

$\bar{x}_1 - \bar{x}_2$ 的標準誤差估計

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

檢定虛無假說 H_0 ：$\mu_1 = \mu_2$ 時：

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$$

$$DF = n_1 + n_2 - 2$$

可信範圍的計算：

$$\bar{x}_1 - \bar{x}_2 \pm t_{\gamma,\ 0.05}\ SE(\bar{x}_1 - \bar{x}_2)$$

$$\gamma = n_1 + n_2 - 2$$

若研究者不滿其結果的不確定性，只有增加樣本數，並重複驗証，因為 n_1 與 n_2 的數目變大時，可消減 $\bar{x}_1 - \bar{x}_2$ 的標準誤差，以得到增加的正確性。

②變異數不相同

若變異數不相等時，可以量度法來解，醫學實驗的分析大都採用此法。若二組平均值與變異數都差異很大時，可以對數值來代替原來的數值，如此，平均值雖仍不相等，但變異數較近似了。若原來兩組平均值差異不大時，便難以改變變異數的差異了。

以上的計算法乃是依據 t 分佈法的特性來考慮，其缺點是變異數爲採用聯合估計值。研究統計資料最好是先計算出二平均值差的標準誤差，即：

$$SE(\bar{x}_1-\bar{x}_2)=\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

虛無假說的檢定在於計算下值：

$$d=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

如果 n_1 及 n_2 足夠大時，上式之值便近似一標準化常態偏量，它的可信範圍：

$$\bar{x}_1-\bar{x}_2\pm Z_{2a}\,SE(\bar{x}_1-\bar{x}_2)$$

Z_{2a} 代表相等於兩邊檢定法時機率爲 2a 的標準化常態偏量。

但若 n_1 及 n_2 之值不夠大時，採用此法則並不精準，因爲上述的 d 值乃是按 S_1^2/S_2^2，n_1 及 n_2 的大小而改變。

兩百分率的比較 Comparison of Two Proportions

依照前述的研討重點，我們可以配對與非配對兩種情況來分析兩百分率的比較。

配對資料 Paired Case

假設兩組樣本各有 N 觀察值，便構成 N 對的觀察值，若設每一分爲爲甲或非甲（如班上同學男爲甲，女爲非甲），這就產生了四種不同的配對：

類型	第一組	第二組	成對數
1	甲	甲	k
2	甲	非甲	r
3	非甲	甲	s
4	非甲	非甲	m

也可以另一方式，按對數排為：

		第二組		
		甲	非甲	
第一組	甲	k	r	k＋r
	非甲	s	m	s＋m
		k＋s	r＋m	N

由此可知在第一組中之甲為（k＋r）/N，第二組中之甲為（k＋s）/N，此二組中甲的百分率之差異為（r－s）/N。

我們先來討論顯著性檢定，也就是虛無假說為（r－s）/N的期望值等於零，也可換句話說是r與s的期望值相等。先將r與s相加是較方便的做法，因為它們具有不相同的二個類型，現用n來代表r＋s，由虛無假中討論，在n相異對中，第2型的對數形態呈二項分佈，而這分佈的參數值為1/2。

當樣本數大時，標準化常態偏量為：

$$2=\frac{r-\frac{1}{2}n}{\frac{1}{2}\sqrt{n}}$$

假如將連續性校正法列入考慮，再將$r-\frac{1}{2}n$的絕對值減去$\frac{1}{2}$再代分分子中，這種檢定法就叫做Mcnemar氏檢定法。

計算二組樣本百分率差異的近似可信範圍時，先求出標準誤差$\sqrt{(r+s)/N}$然後再依一般常態分佈的方法來計算。

雖然顯著性的檢定完全視 r 與 s 的頻數來決定，但其所估百分率和它的標準誤差仍與總數N 有關，也可說是依未配對的來決定有無差異的証明。至於差異的多寡則必須看其餘的資料。

未配對資料Unpaired Case

設二組樣本中，某事件甲發生的機率分別爲 π_1 與 π_2，自第一組中抽取樣本n，當中包含甲的有 r_1，它的比率就是 $P_1 = r_1 / n_1$；再自第二組中抽取樣本 n_2，當中包含甲的有 r_2，其比率爲 $P_2 = r_2 / n_2$。

$$E(P_1 - P_2) = \pi_1 - \pi_2$$

$$\mathrm{var}(P_1 - P_2) = \frac{\pi(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$$

在計算可信範圍時，由於 π_1 與 π_2 是未知數，所以以 P_1 與 P_2 各替代之，卽：

$$\mathrm{var}(p_1 - p_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

$$q = 1 - p_1 \quad ; \quad q_2 = 1 - p_2$$

之後再依據常態分佈的原則，來求出它可信範圍的近似值。

若想檢定虛無假設 H_0：$\pi_1 = \pi_2$，也就是該二組具有一共同 π 值。P_1 與 P_2 都是 π 的最正確估計值。若虛無假設可成立，那麼抽出的二組樣本應是來自同一樣本，所以 π 值的最正確估計值，可將二組樣本合併爲一之後再求出，如此則：

$$P = \frac{r_1 + r_2}{n_1 + n_2}$$

上值爲 π_1 與 π_2 的聯合估計值，得：

$$\mathrm{var}(p_1 - p_2) = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

式中 $q = 1 - p$

故，虛無假說的檢定是計算標準常態偏量爲：

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

雖然使用 p，但是標準誤差的變化很小，若 n_1 及 n_2 不相等，又若 p_1 及 p_2 的差數很大時，它的變化也就很大。其他情況下則可以不同的公式來計算它的標準誤差。

二、統計推論係根據樣本資料以推論或估計全體的各項特性者

通常統計方法可分爲敘述統計（ descriptive statistics ）及統計推論（ statistical inference or statistical estimation ）二種。敘述統計是使用簡單數字表示樣本（ sample ）的各項統計數（ statistics ）。統計推論則依據樣本的統計資料來計算全部資料的各種特點，例如按樣本的標準偏差來求計算全部的標準偏差，或按樣本的平均以估算全部的平均值……等等，皆屬於統計推論的範圍。可見，統計推論是由樣本推論全體，由部份推論全部，由已知推論未知的一種統計方法。

依抽樣方式的不同，樣本的性質也會有所不同，一般是隨機樣本（ random sample ），意思是說全部的一組資料中，每一分子被抽取爲樣本的機會都是相同的，這些被抽取出的樣本的各項估算都足爲全部資料的近似估計，近似值的準確度的多寡視其抽取之樣本的多寡而定，一般認爲，包括 30 項以上的分子便爲大樣本（ large sample ），30 項以下者則爲小樣本（ small sample ），但也不一定，也有人認爲 50 或 100 以上才能稱爲是大樣本，沒有一定標準。樣本數大時，它的分配型式與統計數就越近似全部資料的值數，如果樣本數有 100 項以上的話，它的近似程度就頗高了。

但是，樣本與實際上的全體總會有差距的，儘管樣本數再大，只要不等於全體，則它們之間的差距就可能永遠存在，它們之間差異的大小除了與樣本的大小有關外，散佈地區的不同也是造成差異的原因之一。

樣本平均的顯著性檢定Signiticance lests on A sample Mean

顯著性檢定就是研究樣本的結論其所代表之意義，是統計推論中最爲首要的一種方法。也就是應用統計檢定來分析研究結果，對於待研究的問題，首先設立一個假定，也就是一般人稱的虛無假說（ null hypothesis ），然後依實驗的結果，利用統計分析來驗証這個虛無假說是否能成立，若虛無假說經肯定，我們所研究的結論在統計上便有顯著的意義（ statistically significant ）。如果研究的結果，証實虛無假說不能成立，則此項研究便毫無顯著性的意義。如虛無假說經驗証無法成立，就必須重新設定假說，或甚至改變研究方法，或是增大樣本的數量。

選擇單邊或雙邊的試驗一般是依據問題的性質來決定，如下述：

虛無假設 H_0 ：$\mu=\mu_0$ ，已知Q。

選定 α

H_0 ：$\mu=\mu_0$

假設 $\mu=\mu_0$ ，$\overline{x}$ 爲常態分佈，μ_0 爲平均值，標準偏差爲 $Q_0/\sqrt{n}$

，計算 $Z=\dfrac{\overline{x}-\mu_0}{Q_0/\sqrt{n}}$

①單邊檢定法One - sided test

ⓐ問題：$\mu=\mu_0$ 或 $\mu>\mu_0$ ？

p＝樣本平均值比實際所求得之值大時的百分率

＝Z值右邊面積的大小

若 $p>\alpha$ ，虛無假設 H_0 便可成立（ $\mu=\mu_0$ ）

若 p < α，虛無假設不成立（ $\mu < \mu_0$ ）

ⓑ問題： $\mu = \mu_0$ 或 $\mu < \mu_0$ ？

p ＝樣本平均值比實驗証明的小的百分率

＝ Z 值左方面積的大小

若 p > α，虛無假設成立（ $\mu = \mu_0$ ）

若 p < α，虛無假設不成立（ $\mu < \mu_0$ ）

②雙邊檢定法 Two - sided test

問題： $\mu = \mu_0$ 或 $\mu \neq \mu_0$ ？

p ＝實際求得之樣本平均值與 μ_0 差異之百分率

若 p > α，虛無假設 H_0 成立（ $\mu = \mu_0$ ）

若 p < α，虛無假設 H_0 不成立（ $\mu \neq \mu_0$ ）

t 分佈 The t distribution

實際全體變異數 Q^2 的數值很難確實知道，所以標準平均值的誤差無法求出，$Q/\sqrt{n}$，所以只能用樣本的標準誤差 s 值來推算。S 只是 Q 的近似值，兩者間之差異在所難免，唯樣本數大時，差距較小，樣本數小時，差距就會增大，關於此一環，英國統計學家 W. S. Gosset 氏研究，提出了一個新的變數 t，以代替 z 分佈，後經 R. A. Fisher 研究建立了 t 比例（ t ration ）的抽樣分佈，t 比例基本上與 Gosset 氏所得類似，如下式：

$$t = \frac{(x - \mu_0)}{\frac{s}{\sqrt{n}}}$$

這新變數 t 的分佈不同於 z 變數的常態分佈，它的平均值等於 0，變異數爲 1 。t 分佈的特性是成單峯對稱分佈，其分佈形狀隨自由度（ degree of freedom，d. f., DF ）或 v 而變換（ d. f. ＝ n － 1，即樣本數減 1 ）。若 d. f. 很小時，這個分佈圖形爲高狹峯，長尾踞，若 d. f. 增大時（ 大於 30 ），t 分佈便接近常態分佈。

由於t分佈乃隨著d. f.在變化，基線上某段距離所含曲線下的面積（機率），也隨d. f.而變，所以t比例之機率值乃依據d. f.值來決定。當u＝∞（無限大）時，t分佈與常態分佈相等，也就是，x數值的分佈爲常態時，t分佈的價值則高。

平均值的範圍估計 Interval Estimation of a Mean

在統計學上常使用樣本常數來推算全體參數（parameters）有點推定值（point estimate）與範圍推定值（interval estimate）二種方法。

①點推定值

是依據樣本常數以爲全體常數的最適當推定值，比如全體平均數（μ）應以樣本平均數（$\overline{x}$）來代表最爲合適，則$\overline{x}$爲μ的最佳推定值（the best single estimate）。

②範圍推定值

也叫作信賴範圍（confidence interval）或區間推定。就是以樣本常數去決定全體參數的範圍，優點是，當區間越短，此範圍推定值的正確度就愈高。

選擇信賴──→水準因人而異，一般多以90，95，99％。剛提到區間越短正確度越高，但是太長也不是很理想的。

信賴水準的選擇對各種實驗的正確度影響頗大，如要信賴水準高，又要信賴範圍短，唯一的辦法就是增加樣本。

所需樣本數的事先估算影響實驗很大，決定樣本數要考慮下列數點

①我們所要研究的是什麼？

②期望信賴的水準是什麼？

③預定的信賴範圍。

④全體之變異數。

只要將s代入Q，t值代替z值，便能應用樣本標準偏差來求出全

體平均μ的信賴範圍，當Q爲已知，μ值的95％信賴範圍的計算公式如下：

$$\bar{x} \pm Z - Q/\sqrt{n}$$

因爲$t = \dfrac{\bar{x} - u}{s/\sqrt{n}}$，所以利用s算出$\mu$值的95％信賴範圍，其公式爲：

$$\bar{x} \pm t\gamma，0.05\ (s/\sqrt{n})$$

上公式的意思是，在t分佈中，自由度爲$\gamma = d.\ f. = n - 1$時，t值存在於$\pm t\gamma$，0.05範圍中之機率爲0.95。

百分率推論 Inferences From Proportions

假設在一組資料中抽取樣本，若已知A之機率爲π，抽取n分子，所得A分子的百分率爲P（$= r / n$），其與π之間的關係是什麼？

首先須設立虛無假說H_0：$\pi = \pi$，若在抽取的樣本數n中，所得到A的分子數爲r，是成二項式分佈，爲了解釋任一觀察值r和它的期望值$n\pi_0$之間的差距，可視r爲出現於抽樣分佈曲線之兩端尾踞中。問題是計算偏差值$r - n\pi_0$的機率時，該求大值或小值？由於討論觀察值差異的程度，所以計算機率和才是合理的。

在參數爲π_0及n之二項分佈中，若$r > n\pi_0$，其機率各爲P_0，P_1……，P_n採用單邊檢定法所得的P值應爲

$$P_+ = P_r + P_{r+1} + \cdots\cdots + P_n$$

若應用雙邊檢定法，可將較小部分機率相加，P值爲

$$P_- = P_r{}' + P_r{}'_{-1} + \cdots\ \cdots + P_0$$

而$r' = 2n\pi_0 - r$

應用雙邊檢定法的P值爲$P = P_- + P_+$。

例如：

$r = 17$，$n = 20$，$n_0 = 1/3$

$P_+ = P_{17} + P_{18} + P_{19} + P_{20}$

$P_- = 0$

此外，還有一種方法，就是將單邊檢定時所得到的機率值乘以 2 ，當作雙邊檢定之機率也可以。

最方便的方法就是將兩項分佈當作近似於常態分佈來計算。

按虛無假說，當 $\dfrac{r - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$ 是近似一標準化之常態偏量。用連續性校正法檢定顯著性計算 Z 值

$$Z = \frac{|r - n\pi_0| - \frac{1}{2}}{\sqrt{n\pi_0(1-\pi_0)}}$$

依計算出的機率數值來決定虛無假說的成立與否。

常態近似法有三種：

①從 Z 值計算尾端面積，也就是求 95％的信賴範圍，可套用下二式：

ⓐ $$\frac{r - n\pi_L - \frac{1}{2}}{\sqrt{n\pi_L(1-\pi_L)}} = 1.96$$

ⓑ $$\frac{r - n\pi_V + \frac{1}{2}}{\sqrt{n\pi_V(1-\pi_V)}} = -1.96$$

②如果應用上式之法時，若 n 值很大，則可省略 1/2 的加減計算。此二法皆為求解二次方程式 π_L 和 π_V 。

③用 $P(1-P)$ 代替上二式的 $\pi_L(1-\pi_L)$ 與 $\pi_V(1-\pi_V)$，這是由於 P 值近 1/2 時，$P(1-P)$ 的變化比 P 慢之故。通常求 95％信賴範圍的公式如下：

$$P \pm 1.96\sqrt{pq/n}$$

$q = 1 - p$

三、四重表及卡方檢定法的意義和應用

四重表Fourfold Tables

日　　期	死亡	受傷	合計
上月車禍	5	16	21
本月車禍	8	20	28
總　　計	13	36	49

上面所學的就是四重表的例子，也可叫做2 × 2 列聯表（2 × 2 contingency table）。表中右下角的數是總數，依照兩個橫行數及兩個直行數分開。各個人數分別列於四個表格內，行及列的合計數分別載於右方及下方，這四個數值都叫做邊際和（marginal tatals）。

本節所討論的是非配對的數據（unpaired）與以2 × 2 表的方法來比較樣本中的百分率二者目的小有相異之處。

倘若上月與本月的車禍死亡率與受傷率均相等，我們便能很容易的由邊際和求出死亡人數，這個數值就是期望值（expected）。但今死亡率爲13/49，所以表中左上格中的預估死亡人數爲

$$\frac{13 \times 21}{49} = 5.571$$

應用上例，同樣可求出其他內格中的期望數值，基於方便起見，以O表觀察值（observed），E代表期望值。

期望值並不是眞實確切的，只是一種平均數值，觀察值與期望值之間總會有距離（O－E），但內四格中的O及E差異數值一樣，則爲二正二負。

我們可以這麼說，假如觀察值和期望值二者的差數很大時，虛無假

說就不能成立，虛無假設的能成立與否，端賴顯著性檢定法的檢定，並且將數值的大小列入考慮。

計算指數 x^2 乃是顯著性檢定的驗証法，以公式表示：

$$x^2 = \Sigma \frac{(o - E)^2}{E}$$

期望值越大，便越近似眞值。

x^2 的計算公式種類很多，我們且依上列之表作一公式如下：

	1	2	
1	A	B	r_1
2	C	D	r_2
	S_1	S_2	N

上表之公式為

$$x^2 = \frac{(AD - BC)^2 N}{r_1 r_2 S_1 S_2}$$

顯著性的檢定方法的選擇，須注意以下兩點：

①若應用標準誤方法時，檢定顯著性的有或無不但可檢定出，同時也可求出信賴範圍。照常理，統計分析除了要檢定其顯著性，同時也要能預估由於抽樣誤差所帶來的相關差異，所以在檢定方法的選擇上常計算它的差異與標準誤，而捨棄 x^2 指數的檢定法。

②x^2 檢定法有個上述檢定法所欠缺的優點，就是此法可適用於多種列聯表。採用此法有一點須詳加注意，就是表中的數值一定要是次數才可以，假如數值是變數的平均值就不能應用此法了。

次數的多寡決定這二種方法的是否正確，假如次數不夠大時，必須使用連續性校正法或者是正確機率計算法。

四重表的連續性核對法 Continuity Correction For Four-fold Tables

此法乃係F Yates所提出，所以學者多稱之爲Yates氏校正法（Yatess' correction）。根據虛無假說，$x^2_{(1)}$與x^2的分佈相當接近，所以應視固定的邊際和來決定。基於此局限，合用的次數表也相當有限。爲了要提高四重表中機率的估計值，Yates氏的校正法是先將O－E的絕對值減去$\frac{1}{2}$，再計算x^2，雖然比未校正前少一點，但顯著性仍是相當高的。

期望值如果不大，連續性校正對結果的影響較大，但當以x^2檢定法來分析2×2列表時仍以之處理才好。公式亦要改爲：

$$x_C^2 = \frac{(\ |AD - BC| - \frac{1}{2}N\ )^2\ N}{r_1 r_2 S_1 S_2}$$

連續性的校正法也可應用於標準誤檢定法，但要先將觀察值先減去半個單位後再算$P_1 - P_2$，所得的標準誤不變。

四重表的正確機率計算法The Exact Test For Fourfold Tables

如果次數很小，雖可使用連續性校正法，但是關於x^2的近似性無法肯定。R. A. Fisher、J. O. Irwin和F. Yates等人於1930年代，幾乎同時提出這一正確性檢定法。

以公式$\frac{r_1!\ r_2!\ S_1!\ S_2}{N!\ A!\ B!\ C!\ D!}$來計算下列分佈的正確機率：

A	B	r_1
C	D	r_2
S_1	S_2	N

如果意義水準是10％，仍無顯著性結果，就不該應用單邊檢定法，而改用雙邊檢定法。

依Cochran之提議，使用正確性檢定法和連續性校正法，最好有以下條件：ⓐN＜20 ⓑ20＜N＜40 ⓒ最小的期望值小於5。有這些條

件爲基礎使用正確性檢定法方才適宜。

兩組以上百分率的比較：

$2 \times k$ 列聯表 Comparison of Several Proportions : The $2\times k$ Contingency table

使用 x^2 檢定法與使用 2×2 表相同，對於觀察數O，可用下式求出期望數：

$$E = \frac{\text{行中總數} \times \text{列中總數}}{N}$$

之後，再計算 $(O-E)^2/E$ 的數值，得

$$x^2 = \Sigma \frac{(O-E)^2}{E}$$

研究虛無假說，是假設全部的樣本都是由全體中隨意抽取，各個比例均相等，x^2 的分佈與 $x^2_{(k-1)}$ 相近，它們關係的成立乃視期望數的增加來決定，這時，連續性校正便可省略不用，因爲除非是觀察數很小，而且可將之合併，所以 x^2 的分佈要比 2×2 表更接近連續性分佈。

此外，也可利用另一方法來計算 x^2，就是將表中第一行的數值 $O-E$，x^2 的值則爲 $(ri - P_n i)^2 \left(\frac{1}{P_n i} + \frac{1}{Q_n i}\right)$，式中 $Q = 1 - P$，可簡化成

$$x^2 = \frac{\Sigma_n i\,(Pi - P)^2}{PQ}$$

總合之符號表示包括k組，將上式演算得

$$x^2 = \frac{\Sigma_n i\,Pi^2 - NP^2}{PQ}$$

與

$$x^2 = \frac{\Sigma(r_i^2/ni) - R^2/N}{PQ}$$

兩變異數的比較 Comparison of Two Variarces

設有二獨立樣本，第一組爲 n_1 個樣本數，它的估計變異數是 S_1^2，

第二組爲 n_2 個樣本數，它的估計變異數是 S_2^2 。假使各組樣本都是由常態分佈中抽取，則比較變異數可套用公式求出。這時候個別全體的變異數 Q_1^2 與 Q_2^2 都是未知數。

求信賴範圍的求法，公式中的 Q_1^4 與 Q_2^4 可用樣本估計值 S_1^4 與 S_2^4 來代替，S^4 的替代可決定虛無假說是否能成立。S^2 代表聯合估計變異數 (pooled estimate of variance) :

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

因爲 $S_1^2 - S_2^2$ 的分佈除非在樣本數極大無法接近常態分佈，所以在比較兩變異數的差距時，很少計算它的標準誤差。因此，最好的方法是考慮兩變異數的比值：

$$F = \frac{S_1^2}{S_2^2}$$

若虛無假說成立，自常態分佈中重複抽取樣本 n_1 與 n_2 加以配對，F 分佈的正確數值就可求出。F 分佈的性質是決定於 n_1 與 n_2 的大小值，也決定於自由度($\gamma_1 = n_1 - 1$ ，$\gamma_2 = n_2 - 1$)，但與全體共同的變異數無關。

以 F 值來討論虛無假說時，不論其數值大小，均有顯著性意義的可能，因爲在特定的 γ_1 與 γ_2 ，F 的數值可大於 1 ，也可小於 1 。表列的全部臨界值 (critical value) 都是大於 1 ，所以只應用 F 分佈的上端。由於以 1 和 2 來表示樣本是隨意抽取的，所以查附表並無不便之處，只要 1 與 2 互相異位，比值就可由小於 1 變成大於 1 。

兩變異數中的大值以 S_1^2 表示，以 γ_1 表其自由度。如果要利用雙邊檢定法，要將表中之 P 值減半以成意義水準。

用 F 的分佈表亦可求出 Q_1^2 / Q_2^2 的信賴範圍，假使虛無假說不成立，$F = \frac{S_1^2 / Q_1^2}{S_2^2 / Q_2^2}$ 應符合 F 分佈，並且配合 γ_1 與 γ_2 的自由度(虛無假

說是 $Q_1^2 = Q_2^2$，$F' = F$ ），這種情況下就不以 $S_1^2 > S_2^2$ 來表示。用符號 $F\alpha$，γ_1，γ_2 是代表自由度爲 γ_1，γ_2 時，單邊意義水準 α 的條件下所得的F臨界值，所以用 $F\alpha$，γ_1，γ_2 表示時，它的意義同於上述，僅有自由度 γ_1 與 γ_2 互換罷了。這時候機率爲 α，

$$F' > F\alpha, \gamma_1, \gamma_2$$

也就是 $F > F\alpha, \gamma_1, \gamma_2 \, (Q_1^2 / Q_2^2)$

$$1 / F' > F\alpha, \gamma_2, \gamma_1$$

也就是 $F > (1 / F\alpha, \gamma_2, \gamma_1) (Q_1^2 / Q_2^2)$

機率是 $1 - 2\alpha$

$(1 / F\alpha, \gamma_2, \gamma_1) (Q_1^2 / Q_2^2) < F < F\alpha, \gamma_1, \gamma_2 (Q_1^2 / Q_2^2)$ 可化簡成：

$$F / F\alpha, \gamma_1, \gamma_2 < Q_1^2 / Q_2^2 < F / 1\alpha, \gamma_2, \gamma_1$$

F檢定法與它的信賴範圍的計算用途是用來比較二個不同的常態樣本中的變異數。它的缺點是對於常態性的假設太過於敏感，也就是健全性不夠，比t分佈的平均值比較尤爲甚者。

F檢定法用來比較二組獨立變數較適宜，若是用來比較已經過配對後的樣本，有時是解不出的。

一般列聯表（Generel Contigency Tables）

在組數百分率的比較中的表式可一般化，假設總次數是N，然後將它分成r行及c列項數。虛無假說是設各列中的發生機率與行數無關，或者各行中的機率與各列的一樣。

檢定 x^2 是先將公式

$$E = \frac{\text{行中總數} \times \text{列中總數}}{N}$$

求出 $(O - E)^2 / E$ 的數值，便得

$$x^2 = \Sigma \frac{(O-E)^2}{E}$$

求出各表格中的期望數E，然後再相加 r_0 格內的 x^2 值即可，虛無假設 x^2 的分佈符合 $x^2_{(f)}$ 的分佈，自由度 $f=(r-1)(c-1)$。

Cochran 氏認爲，雖然 $x^2_{(f)}$ 分佈在期望數增大時會越有它的近似性，但期望數若大於5時就沒有問題，並且無一期望數小於1。所以，當期望數小的時候，如果顯著性的意義不明顯，可將數行或數列中之數值重新予以合併，以重新求算 x^2。

第六章
差異量數的分類與意義

一、標準差爲一種極重要而且又有價值的差異量數

標準差是一種最重要且最具有統計上的價值，又可稱爲均方差，是各量數和算術平均數的差數平方和平均的平方根。但由於絕對值祇爲計算根據，不適宜代數方式處理，補救辦法就是利用標準差的計算，先計算出差數的平方和，然後再加以平均，最後的平方根是一還原的步驟，所以，標準差既較全距與四分位差反應靈敏，且又比平均差爲佳，所以應用於統計較爲多。

$$S=\sqrt{\frac{\Sigma(x-M)^2}{N}}$$

公式〔6－1〕

$$S=\sqrt{\frac{\Sigma x^2}{N}-\left(\frac{\Sigma x}{N}\right)^2}$$

$$=\frac{1}{N}\sqrt{N\Sigma x^2-(\Sigma x)^2}$$

公式〔6－2〕

$$S=\frac{1}{N}\sqrt{N\Sigma x^2-(\Sigma x)^2}$$

公式〔6－3〕

$$S = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

公式〔6－4〕

S：標準差（常態曲線中的標準差是以Q表示）

以上四個公式都是量數未分類時，計算標準差的公式。

公式〔6－1〕是先求算術平均數之後再求標準差。公式〔6－2〕是由原來的變量計算標準差。公式〔6－3〕與公式〔6－2〕相同，只是將之簡化而已。也是由原來變量求標準差。公式〔6－3〕中的 $x = x - x'$，x' 是依據實際情況而定的一定數值，各變量 x 減去相同數值 x' 之後，數值變小，方便計算。公式〔6－4〕是以假定平均數求標準差，其中的 $d = x - A$，由上述四個公式求出的標準差均會相同。

若數量已經歸類分組，可以下列公式計算標準差：

$$S = \sqrt{\frac{\Sigma f(x - M)^2}{N}}$$

公式〔6－5〕

$$S = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$$

公式〔6－6〕

如果各組組距相同，公式〔6－6〕便可改寫成：

$$S = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \quad i$$

公式〔6－7〕

N：代表總次數

f：各組次數

x：各組中點

M：算術平均數

d：x－A

A：假定平均數

i：組距

d′：$\frac{x-A}{i}$

從已分組歸類的計算標準差資料，是假設各組的量數都集中在組距中點，並以各組中點為計算根據，但是這個假設和各變量的實際分佈情形，不容易完全符合，所以，從次數分配表計算出的標準差跟從原來未分組歸類時所計算出的結果都不大相同。由分組而呈現的誤差，隨著組距的變化而變化，假使資料是屬連續變量且為單峯分配，並近似常態鐘形的情形時，可以謝巴德（V. F. Sheppard）校正公式來減少分組的誤差。

$$〔S〕=\sqrt{\frac{\Sigma fd'^2}{N}-(\frac{\Sigma fd'}{N})^2-\frac{1}{12}}\ i$$

公式〔6－8〕

也可寫為：

$$〔S〕=\frac{i}{N}\sqrt{N\Sigma fd'^2-(\Sigma fd')^2-0.0833N^2}$$

公式〔6－9〕

若已求出分組資料的標準差，就可將上述公式改寫為：

$$〔S〕=\sqrt{S^2-\frac{i^2}{12}}=\sqrt{S^2-0.0833i^2}$$

公式〔6－10〕

〔S〕：校正後的標準差

S：由分組資料求出的標準差

其餘符號與前述同

若分組資料有十組以上，用一般普通的公式求標準差就可以了，不須要應用到校正公式。一般在鐘形單峯分配中，以各組中點為準將組內變量一分為二時，接近平均數的一部分要比平均數一半的項數多得多。

所以以組中點做為組內各項的代表值時，往往會誇大各數值與平均數之間的差距。標準差的計算，是以該數加以平方，所以，自分組資料所計算出的標準差可能過大，所以須要校正。但有一點要注意的是，若分組資料不是單峯接近鐘形分配與非連續變量，便不能使用校正公式。

標準差的特點

①是以算術平均數為中點所求出的標準差，比其他數為中點所求出的結果小。

②標準差的正確性要比平均差優越，因為標準差是依據全體量數計算而得。

③標準差受抽樣變動的影響少。

④標準差適合代數方法處理，部分資料的項數、算術平均數及標準差為已知數，可以合併求出全體資料的標準差：

$$S=\sqrt{\frac{N_1(S_1^2+M_1^2)+N_2(S_2^2+M_2^2)\cdots N_k(S_k^2+M_k^2)-M^2}{N_1+N_2+\cdots+N_k}}$$

公式〔6－11〕

S：全體資料的標準差

M：全體資料的算術平均數

N_1，$N_2\cdots N_k$：是表示各部分資料的項數

M_1，M_2，$\cdots M_k$：是代表各個部分資料的算術平均數

S_1，S_2，$\cdots S_k$：各部分資料的標準差

標準差在次數圖的橫座標上是一段部分，於常態分配上，從M向左方或右方劃一標準差的距離，包括全部曲線下的34.13％面積，兩個標準差的距離又包括47.72％的面積，三個標準差的距離又包括49.87％的面積。也就是由M左右各量一個標準差的距離範圍，共包括全部面積的68.26％。一般常態分配的底線，大多以標量差為量尺單位，照說這個底線應是無限長的，但在M的上下方各量三個標準差（也就等於六個

標準差的長度），就已經包括了全部面積的 99.7 %以上了。

使用標準差可使原來的量數化爲標準量數，可直接比較原本單位不一的量數，也可相加、減或平均，標準量數的計算公式如下：

$$Z = \frac{X - M}{S}$$

公式〔 6 － 12 〕

Z ：代表標準量數

其餘符號與前述同

研究公式〔 6 － 12 〕可知標準量數的 0 就是算術平均數，標準量數是 1 ，也就是在算術平均數上的一個標準差的位置，標準量數－ 1 ，也就是在算術平均數之下的一個標準差的位置。

標準是各項差異量數之最佳者，它除了受限於極端量數的影響與分組資料中有不確定組距時不能求出標準差外，在統計上是應用最爲普遍的，常和算術平均數一起計算，若算術平均數是適當的集中量數時，標準差就是最適當的差異量數。

爲了避免計算上的不便，當標準量數可能爲小數或負數時，通常都先乘以常數 10 ，再加上常數 50 ，如此一來，便可改寫公式〔 6 － 12 〕爲：

$$Z' = 10Z + 50$$

公式〔 6 － 13 〕

Z ：原來標準量數

Z′：轉換後的標準量數

均互差

一群數值中，每兩個數值相差絕對值總和的算術平均數就是所謂的相互差，是意大利統計學家 Corrods Gini 所提出的差異量數，一般都以 g 表示之。

設有N個數值，其中每兩個數值的相互差數共有$\frac{N(N-1)}{2}$個，也就是它的組合數爲$_{N}C_{2}$個，若項數N多，組合便多，使用一般的普通方法計算較不方便，可依據資料的分組與否採用簡捷法來計算：

$$g=\frac{(N+1)\Sigma X-2Sa}{\frac{N(N-1)}{2}}$$

公式〔6－14〕

公式〔6－14〕是未分組資料計算相互差的簡捷法，

g：相互差

N：項數

ΣX：N個數值的總和

Sa：N個數值由小到大依順序排列的累積數的總和

假使量數已經歸類分組，且組距相同，則可應用下述公式計算：

$$g=\frac{\Sigma F(N-F)}{\frac{N(N-1)}{2}}\times i$$

公式〔6－15〕

F：各組的累積次數

其餘各符號與上同

在統計上很少用到均互差，若項數多，資料尚未分組，計算上較不方便，而且分組資料以組中點代表而忽略了組內各數值間的差量，難符合原來的意義，所以很少應用到。

相對差數

一般常用的相對差數是皮爾生（K. Pearson）所創的，公式如下

$$V=\frac{S}{M}\times 100$$

公式〔6－16〕

V：相對差數

其餘符號與前述同

它的優點是，使原有差數變成抽象數值，少了單位的局限，可比較不同單位的差距，並也可減少平均數大小不一的影響。

前述的全距、四分位差、標準差、平均差與均互差都是具有與原來資料相同的單位，是屬有名數，都叫做絕對差數 Absolute Variability。如果兩組資料的單位不一樣時便無法使用絕對差數來直接比較差異的大小。若平均數的大小相差太大，即使單位相同，也無法直接比較。爲了補救此遺漏，統計學上便利用分配中的差異量數與有關平均數的比率作爲比較各種差異程度的標準，叫做相對差 Relative Variability 或差異係數（Coefficient of Variation）。

除了公式〔6－16〕是用同一資料的標準差和算術平均數計算相對差外，其他的差異量數也可以同樣方法求相對差數。例如全距和最大最小兩數的和或者算術平均數的比值，四分位差和中位數或上及下兩個四分點和數平均的比值，相互差與算術平均數或中位數的比值，都是相對差數。

二、差異量數大表示各變量比較參差不齊

一組變量彼此差異分佈的情形就是差異量數，是次數分配的特徵之一，於次數分配圖中是以距離來表示，所以也叫做距離量數（Distance measures）。若各變異量彼此相近，則差異量數小，變化較小。差異數量的大小，除決定變異程度的大小外，同時亦可決定平均數的價值程度，平均數代表價值的程度，與差異量數的大小成相反的方向。

一般常用的差異量數，除了上節中討論過的標準差與相互差外，尙有全距、四分差及平均差三者，五五種，爲了比較兩種以上統計資料的差異程度，因單位的不一，平均數大小也不相等的因素，不容易使用上述五種差異量數來直接比較，因而產生了差異係數（Coefficient

of Variation)。

現將全距、四分差及平均差三者，分別詳述於下：

全距

在所有的變量中，最大數與最小數的差距就是全距。大多數的人認為，在分組資料中求全距，應以組值最大的高限減去組值最小的低限之後所得到的差距就是全距。如此所得之全距，比起未分組前為高，因此又有人主張，以組值最大的組中點，減去組值最小的組中點的差距就是全距，但是，如此所得的全距又比起未分組前為低了。折衷的方法就是，以組值最大的組高限減去組值最小的組中點，或者，以組值最大的組中點減去組值最小的組低限，這樣一來就較接近未分組前的全距了。

全距又名兩極差，因為它是將兩極端的大小量數相減而求出的。全距受抽樣變動的影響很大，很不穩定，所以在使用上也就相當有限了，只能當作是離差的補助量數，應少用。

為了避免受抽樣變動、兩極量數差異太大的影響，有人認為應以百分位來替代全距，就是用 $P_{90} - P_{10}$ 的差數為離差。

四分位差

所謂四分位差就是，四分位數中的第三四分位數(Q_3)和第一四分位數(Q_1)間距離的一半。也可說是百分位數中的第七十五百分位數與第二十五百分位數距離的一半。

因為 Q_1 和 Q_3 之間的距離包含了全體量數中中點部份的50%的量數，因此四分位差可以免受兩極端量數的影響。

$$Q = \frac{Q_3 - Q_1}{2}$$

公式〔 6 − 17 〕

Q：四分位差

在單峯對稱的分配圖中，Q_1 及 Q_3 與中位數的距離是相等的，也就

是$Q_3-M_e=M_e-Q_1$，但在偏斜的分配中，二數就不能相解了。

四分位差的使用通常都與中位數相互配合，若以四分位表示差異的情形時，則採用中位數爲集中量數。雖然四分位差比起全距，穩定性較高，但因爲四分位數只是自中間50％的量數計算而得，並未考慮到另外50％的量數的差異，所以還是屬於感應不靈的一種量數。

平均差

平均差乃係一數列中各量數及其中位數或算術平均數之差數絕對值總和的平均數。以差異的觀點來看，無論正差或負差同樣是表示差異，是一樣重要的，況，計算平均差如果不以絕對值爲基準，各正差與負差會相互抵銷，所得結果就可能是0或近0，無法顯示出眞正的差異，這就是爲什麼計算平均差要應用絕對值的原因所在。

因爲以中位數爲準的平均差數值最小，以前的統計多利用中位數求平均差，但由於算術平均數是最常使用的平均數，所以最近的統計學者大多主張以算術平均數來求平均差。

自中位數求平均差的公式：

$$M.D.=\frac{\Sigma|x-M_e|}{N}$$

公式〔6－18〕

由算術平均數求平均差的公式：

$$M.D.=\frac{\Sigma|x-M|}{N}$$

公式〔6－19〕

M.D.：平均差

｜　｜：絕對值

其餘符號與前述同

若量數已分類，可以下列公式求平均差：

$$\mathrm{M.D.} = \frac{\Sigma f\,|\,x - M_e\,|}{N}$$

公式〔 6 － 20 〕

也可寫成：

$$\mathrm{M.D.} = \frac{\Sigma f\,|\,x - M\,|}{N}$$

公式〔 6 － 21 〕

公式〔 6 － 20 〕與公式〔 6 － 21 〕來計算分組資料的平均差不是很方便，如果應用假定平均數來計算平均差會方便許多：

$$\mathrm{M.D.} = \frac{\Sigma\,|\,fd\,| + C\,(\,Fe - Fg\,)}{N}$$

公式〔 6 － 22 〕

d ：各組中點與假定平均數（ A ）的差數

C ：眞正平均數與假定平均數的差數

Fe ：組中點小於眞正平均數的各組的次數和

Fg ：組中點大於眞正平均數的各組次數和

其餘符號與前述同

若各組組距相同，可使用下列公式計算平均差：

$$\mathrm{M.D.} = \frac{\Sigma\,|\,fd'\,| + C'\,(\,Fe - Fg\,)}{N}\,i$$

公式〔 6 － 23 〕

d′：各組中點及假定平均數（ A ）的差距，以組距為單位者

C′：眞正平均數與假定平均數的差，以組距為單位者

即：

$$d' = \frac{x - A}{i}$$

$$C' = \frac{Me - A}{i}$$

或者

$$C' = \frac{M - A}{i}$$

假定平均數應以眞正平均數所在組的組中點做爲假定平均數，所以，不管使用的計算公式爲何，都要先計算出中位數或算術平均數。

一般以上列公式求平均差的值已頗近似眞値，但若須求得更精準的結果，可依均勻分佈的假定予以校正，校正公式如下：

$$\frac{f}{N}\left(\frac{C^2}{i} + \frac{i}{4} - |C|\right)$$

公式〔 6 － 24 〕

f ：平均數所在組的次數，也就是Me 或M 所在組的次數

N ：總次數

i ：平均數所在組的組距

C ：平均數與其所在組組中點之差

若各組組距相同，則公式〔 6 － 24 〕可改爲：

$$\frac{fi}{N}\left(C'^2 + \frac{1}{4} - |C'|\right.$$

公式〔 6 － 25 〕

C′：等於$\frac{C}{i}$

因校正公式所得的平均數一定爲正數，所以自公式〔 6 － 20 〕到公式〔 6 －23 〕所得的平均差加上校正後的數值，一定也比原式所計算的結果稍大。

若把公式〔 6 － 25 〕與公式〔 6 － 23 〕二者合用，即可得一計算結果更爲眞確的平均差，以公式如下：

$$M.D.=\frac{i}{N}\left[\Sigma|fd'|+C'(Fe'-Fg')+f\left(C'^2+\frac{1}{4}\right)\right]$$

公式〔6－26〕

Fe′：組中點小於Me或M所在組其他各組的次數和

（不包括Me及M所在組的次數）

Fg′：組中點大於Me或M所在組其他各組的次數和

（不包括Me及M所在組的次數）

f　：Me或M所在組的次數

其他符號與前述同

因爲平均差可自中位數計算，也可由算術平均數計算，所以一般自中位數計算出的平均差與中位數合併使用，由算術平均數所求得的平均差則和算術平均數合併使用。

因平均差是依據全體的變量計算而得，所以可表示數列中全部項目的差異情形，比全距及四分位差靈敏。但平均差也有二個限制：

①易爲極端量數影響。

②計算時使用絕對值，無法分出正負，不符合數理的計算原則。

三、動差的意義及計算

動差

動差是皮爾生尺所提出，是以任何數值爲中心的一級、二級、三級、四級動差，公式如下：

$$m=\frac{\Sigma fd}{N}$$

$$m'2=\frac{\Sigma fd^2}{N}$$

$$m'3 = \frac{\Sigma fd^3}{N}$$

$$m'4 = \frac{\Sigma fd^4}{N}$$

公式〔6－27〕

計算以算術平均數爲中心的一級、二級、三級、四級動差，公式如下：

$$m_1 = \frac{\Sigma fx}{N}$$

$$m_2 = \frac{\Sigma fx^2}{N}$$

$$m_3 = \frac{\Sigma fx^3}{N}$$

$$m_4 = \frac{\Sigma fx^4}{N}$$

公式〔6－28〕

這二種動差的關係是

$$m_1 = \frac{\Sigma fd}{N} - \frac{\Sigma fd}{N}$$

$$= m_1' - m_1'$$

$$= 0$$

$$m_2 = \frac{\Sigma fd^2}{N} - (\frac{\Sigma fd}{N})^2$$

$$m_2' = m_2' - m_1'$$

$$m_3' = \frac{\Sigma fd^3}{N} - 3\frac{\Sigma fd}{N}\frac{\Sigma fd^2}{N} + 2(\frac{\Sigma fd}{N})^3$$

$$m_3 = m'_3 - 3m'_1 m_2 + 2m'^3_1$$

$$m_4 = \frac{\Sigma fd^4}{N} - 4\frac{\Sigma fd}{N}\frac{\Sigma fd^3}{N} + 6\left(\frac{\Sigma fd}{N}\right)^2\left(\frac{\Sigma fd^2}{N}\right) - 3\left(\frac{\Sigma fd}{N}\right)^4$$

$$m_4 = m'_4 - 4m'_1 m'_3 + 6m'^2_1 m'_2 - 3m'^4_1$$

m_1 是以原點（0點）爲中心的一級動差，S^2 是以M爲中心的二級動差，偏態係數是 m_3 與 S^3 之比，峯度係數是 m_4 與 S^4 之比。

四、優良差異量數的條件及準確度

差異量數的比較

各差異量數間的關係

在常態分配或近似常態分配的單峯偏微分配中

O（標準差）＝1.253 M. D（平均差）

O＝1.483Q（四分差）

M. D＝0.798O

M. D＝1.184Q

Q＝0.675O

Q＝0.845 M. D

常態分配或近似常態分配，六倍標準差的距離，七倍半平均差的距離，八倍四分差的距離，大抵可含括百分之九十九的全體量數。

同一資料所得的離中差數

$g > S > M.D. > Q$

優良差異量數的條件：

①計算容易。

②適合代數方法處理。

③受抽樣變動的影響不大。

④容易使人了解。

⑤感應靈敏。

⑥不容易受極端量數的影響。

上述的各差異量數均各有其優點與缺點，一般說來，標準差最佳，應用於統計上最爲普遍，其次爲平均差與均互差，再其次爲四分位差，全距的使用範圍較爲狹窄，只能作補助之用。此外，在不同的需要下，要依其性質及適用的量數使用適當的差異量數。

差異量數的準確度

四分位差因是由點計而得，照理說應該是差異量數中最爲精確的，但由於它是依據 Q_1 及 Q_2 所計算，所以計算結果多爲近似值，而非精確值，Q 自然也多半爲近似值。平均差的公式是 $\frac{\Sigma|X-M|}{N}$ 或 $\frac{\Sigma|X-Me|}{N}$，標準差的公式是 $\sqrt{\frac{\Sigma(x-M)^2}{N}}$，均互差的公式是 $\frac{(N+1)\Sigma X-2Sa}{\frac{N(N-1)}{2}}$，那麼平均差與 $\Sigma|X-M|$ 或 $\Sigma|X-Me|$ 有相等的有效數字，標準差與 $\Sigma|X-M|^2$ 有相等多的有效數字，均互差與 ΣX 有相等多的有效數字

計算分組資料時，不容易求出上述有效數字的位數，爲求統計上的方便，一般差異量數多和平均數有相同的準確度。

偏態係數

各種次數分配的性質，可於下列四種角度來表示：

①可用平均數表示次數分配的中心位置。

②可用差異數量表示次數分配的離中程度。

③可用偏態係數表示次數分配的高峯位置是否居中。

④可用峯度係數表示次數分配的高峯起伏狀態。

所以任何次數分配資料都有平均數、差異量數、偏態係數及峯度係數四種常數，前二者較後二者常用，也較重要。

由算術平均數、中數及衆數三者中可看出次數分配的高峯是否居中

心位置。在對稱的次數分配中，三者合一，若次數分配有所偏斜即三者分離。當次數分配向左偏時，叫做負偏，是$M < M_0$。當次數分配向右偏時，叫做正偏，是$M > M_0$。偏斜度越大，差異也就越大。

因爲偏態可表示離中差的分配情形並爲了使各種次數分配的偏斜情形便於比較，應讓M與M_0的差數變爲抽象量數，成爲一無單位的量數，所以要除以離中差數，使之成爲偏態係數（Coefficient of Skewness）。

一般最爲普遍爲人採用的偏斜係數是K. Pearson所創的，根據M及M_0來計算，公式如下：

$$S.K. = \frac{M - M_0}{S}$$

公式〔6－29〕

S.K.：偏態係數

其餘符號同前述

公式〔6－29〕能將偏斜的方向及偏斜的程度計算出來，廣爲人用，但有三個缺點：

①理論上，變動範圍無限制，對稱分配時爲0，在偏斜不大的單峯分配中則爲＋1與－1之間。

②不容易確定衆數的值。

③不適合代數方法處理。

因爲偏微時$M_0 = M - 3(M - Me)$所以有時可改爲：

$$S.K. = \frac{3(M - Me)}{S}$$

公式〔6－30〕

除了可以M、M_0與Me的數值來計算偏斜係數外，還有可計算偏斜係數的其他公式：

$$S.K.=\frac{(Q_3-Me)-(Me-Q_1)}{Q_3-Q_1}$$

$$=\frac{Q_3+Q_1-2Me}{Q_3-Q_1}$$

公式〔6－31〕

公式〔6－31〕是由Bowley與Yule共同提出，對稱分配時，$Q_3-Me=Me-Q_1$，S. K.便＝0，若為偏斜分配，當$(Q_3-Me)>(Me-Q_1)$時為正偏，當$(Q_3-Me)<(Me-Q_1)$時為負偏。當$Q_3=Me$時，S. K.＝－1，當$Me=Q_1$時，S.K.＝1，所以S. K.的變動範圍在＋1與－1之間。

Bowley認為S. K.＝±0.1時為微偏，S. K.＝±0.3時就算顯著的偏微。這個公式的缺點是敏感度不夠，且不符合代數方法處理。

$$S.K.=\frac{(P_{90}-Me)-(Me-P_{10})}{P_{90}-P_{10}}$$

$$=\frac{P_{90}+P_{10}-2Me}{P_{90}-P_{10}}$$

公式〔6－32〕

公式〔6－32〕在對稱分配時，$P_{90}-Me=Me-P_{10}$，S. K.為0。不對稱分配時，$P_{90}-Me\neq Me-P_{10}$，則係數介於＋1與－1之間，這個公式是M. W. Tate所創立，缺點與公式〔6－31〕類似，反應不夠靈敏，且不適合代數演算。

上述的各項公式的共同缺點都是敏感度不夠及不適合代數演算，下式可計算出較正確的偏態係數：

$$S.K.=\frac{\frac{1}{N}\Sigma(x-M)^3}{S^3}=\frac{\frac{1}{N}\Sigma x^3}{S^3}$$

公式〔6－33〕

若資類已分組，則可改爲：

$$S.K.=\frac{\frac{1}{N}\Sigma f(x-M)^3}{S^3}$$

$$=\frac{\frac{1}{N}\Sigma f x^3}{S^3}$$

公式〔 6 － 34 〕

對稱分配時，Σx^3 或 $\Sigma f x^3$ 係數爲 0，S. K. ＝0，正偏時 S. K. ＞0，負偏時 S. K. ＜0。此公式的優點是：

①適合代數方法處理。

②敏感度高。

③可顯示出偏斜方向與偏斜程度。

缺點是 S. K. 的變動無一定局限。

Croxton 與 Cowden 認爲，若 S. K. ＝± 2 時，表示偏態顯著。

如果M不等於整數時，以上公式的計算就太爲繁複，因此可將公式〔 6 － 34 〕化爲：

$$S.K.=\frac{\frac{\Sigma fd^3}{N}-3\frac{\Sigma fd}{N}\frac{\Sigma fd^2}{N}+2\left(\frac{\Sigma fd}{N}\right)^3}{S^3}$$

公式〔 6 － 35 〕

$d=x-A$

其餘符號與前述同

若組距相同，計算上就更爲便利：

$$S.K.=\frac{\frac{\Sigma fd'^3}{N}-3\frac{\Sigma fd'}{N}\frac{\Sigma fd'^2}{N}+2\left(\frac{\Sigma fd'}{N}\right)^3}{S'^3}$$

公式〔 6 － 36 〕

d' 與 S' 均爲以組距爲單位者。

峯度係數（Measures of Kurtosis）

單峯的次數分配圖中，若項目分配均勻，峯的形狀是平鐘形，稱爲常態峯（Mesokurtic），若項目的分配集中在平均數，則峯狀高窄，稱爲高狹峯（Leptokurtic）。若峯形無明顯的集中趨向，則峯狀成低而濶，稱爲低濶峯（Platykurtic）。

顯示峯的高低程度，要使用無單位的峯度係數，以常態分配的峯度爲基準分爲高狹峯、常態峯與低濶峯三者。

峯的係數可分爲粗略峯度係數及精密峯度係數兩種，粗略峯的計算公式爲：

$$Ku = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$$

公式〔6－37〕

Ku：非代數性的粗略峯係數

此公式是利用次數分配的中間部份50％的數值距離及中間部份80％的數值距離之比例爲峯態係數。如依公式〔6－37〕所計算出的峯態係數是0.263。若分配次數比常態峯密集在中間部份，分子變小，$Ku < 0.263$，卽高狹峯。若次數分配中間部分比常態峯分散，則分子擴大，$Ku > 0.263$，便是低濶峯。

公式〔6－37〕所計算求出的粗略峯係數，並不合適代數處理，僅能做爲量數說明之用，不能以樣本的粗略峯度係數來表示全體的常態狀況，所以在使用上受到相當的局限。

$$K = \frac{\frac{1}{N}\Sigma(x - M)^4}{S^4}$$

$$= \frac{\frac{1}{N}\Sigma x^4}{S^4}$$

公式〔6－38〕

$$K=\frac{\frac{1}{N}\Sigma f(x-M)^4}{S^4}$$

$$=\frac{\frac{1}{N}\Sigma fx^4}{S^4}$$

公式〔 6 － 39 〕

公式〔 6 －38 〕是求未分組資料的精密峯度係數（ K ）的一般使用公式；公式〔 6 －39 〕是使用於已分組資料者。若M不等於整數，演算上極爲不便，此時可改爲：

$$K=\frac{\frac{\Sigma fd^4}{N}-4\frac{\Sigma fd}{N}\frac{\Sigma fd^3}{N}+6\left(\frac{\Sigma fd}{N}\right)^2\frac{\Sigma fd^2}{N}}{S^4}$$

$$-3\left(\frac{\Sigma fd}{N}\right)^4$$

公式〔 6 － 40 〕

d ：x － A

其餘符號與前述同

若已分組資料的組距相同，並以組距爲單位時，可用下式計算較方便：

$$K=\frac{\frac{\Sigma fd'^4}{N}-4\frac{\Sigma fd'}{N}\frac{\Sigma fd'^3}{N}+6\left(\frac{\Sigma fd'}{N}\right)^2\frac{\Sigma fd'^2}{N}}{S'^4}$$

$$-3\left(\frac{\Sigma fd'}{N}\right)^4$$

公式〔 6 － 41 〕

d′與S′都是以組距爲單位者。

第七章

機率爲闡釋統計理論及統計決策的工具

一、常態分佈

常態分佈Normal　Distribution

Poisson 氏分佈與二項分佈皆與間斷無分佈有關。連續型變數的分佈曲線種類很多，各個全體皆不相同。

「常態」（ Normal ）此一名詞與醫學上的Normal「正常」意義並不相同，爲區分先見，學者多稱之爲「高斯氏分佈」（ Gaussian distribution ），以免二者混淆。

常態曲線是呈左右對稱，並且爲吊鐘狀分佈，其方程式如下：

$$f(x) = Y = \frac{N}{\sqrt{2\pi\theta}} e^{-\frac{(x-\mu)^2}{2\theta^2}}$$

$$(-\infty < x < \infty)$$

$f(x)$ ＝機率密度

x ＝隨機變數

Y ＝頻數

μ ＝平均值

θ ＝標準偏差

$\pi \fallingdotseq 3.1416$（圓周率）

$e \fallingdotseq 2.7183$（自然對數之底）

N ＝觀察總數

為了方便演算，通常將常態分佈曲線與橫軸之間的面積視為 1 單位，而不計算它的總面積，所以可簡化上式為：

$$f(x)=Y=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-\mu)^2}{2\theta^2}}$$

常態分佈曲線的特點如下：

①平均值在中點，因其為左右對稱，$\bar{x}=\mu$。

②常態曲線的兩端為無限延伸，但 3θ 以上的部分很小，故常態曲線的部分只演算到 $\mu\pm3\theta$ 也就相差無幾了。

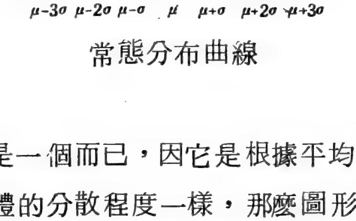
ⓐ $\mu\pm3\theta$ 大抵佔了曲線下的所有面積（99.73％）。

ⓑ $\mu\pm2\theta$ 約佔曲線下的95％面積（95.44％）。

ⓒ $\mu\pm\theta$ 約佔曲線下面積的 2/3（68.26％）。

③自中心點觀之，其為左右對稱。

假設曲線與橫軸間的面積為一單位，則可以圖示之：



常態分布曲線

常態曲線並不只是一個而已，因它是根據平均值及標準偏差的數值來繪的曲線，若兩全體的分散程度一樣，那麼圖形的形狀便大同小異。

若 μ 與 θ 的數值已知，便可繪製常態分佈圖，而且只有一個常態曲線。但是若一數據不知道它是呈常態分佈，即使 μ 與 θ 為已知，亦無法繪出確切的圖形。

標準常識曲線Standard normal curve

常態曲線由上述可知是屬於理論的次數分佈曲線，標準偏差θ是常態曲線的差異數，平均值μ常位於常態曲線的中心，整個曲線下所包含的曲線等於1。

常態分佈的特性：

	1	2	3
變數	x	$x-\mu$	$\frac{x-\mu}{\theta}$
平均值	μ	0	0
標準偏差	θ	θ	1

若1組全體爲常態分佈，第2、3組也都會呈常態分佈，三組不同全體的平均值與標準偏差的演算方式：

①變數＝x

$$平均值=\frac{\Sigma x}{N}=\mu$$

$$標準偏差=\sqrt{\frac{\Sigma(x-\mu)^2}{N}}=\theta$$

②變數＝$x-\mu$

$$平均值=\frac{\Sigma(x-\mu)}{N}=\frac{\Sigma X}{N}-\frac{\Sigma\mu}{N}$$

$$=\mu-\frac{N\mu}{N}=\mu-\mu=0$$

$$標準偏差=\sqrt{\frac{\Sigma[(x-\mu)-0]^2}{N}}$$

$$=\sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

$$=\theta$$

③變數$=\dfrac{x-\mu}{\theta}$

$$平均值=\frac{\Sigma\left(\frac{x-\mu}{\theta}\right)}{N}=\frac{\Sigma(x-\mu)}{N\cdot\Sigma\theta}$$

$$=\frac{\Sigma x}{N\cdot\Sigma\theta}-\frac{\Sigma\mu}{N\Sigma\theta}$$

$$=\frac{\mu}{\Sigma\theta}-\frac{N\mu}{N\Sigma\theta}$$

$$=\frac{\mu}{\Sigma\theta}-\frac{\mu}{\Sigma\theta}$$

$$=0$$

$$標準偏差=\sqrt{\frac{\Sigma\left[\left(\frac{x-\mu}{\theta}\right)-0\right]^2}{N}}$$

$$=\sqrt{\frac{\Sigma\left(\frac{x-\mu}{\theta}\right)^2}{N}}$$

$$=\sqrt{\frac{\Sigma(x-\mu)^2}{N\cdot\theta^2}}$$

$$=\frac{\theta}{\theta}$$

$$=1$$

第3組全體的常態分佈，平均值$\mu=0$，標準偏差$\theta=1$，我們稱此種全體的變數爲標準常態變數（ standard normal variate ）或是相對偏量（ relative deviate ），或臨界比值（ Critical ration ）或正常偏量（ normal deviate ）或標準比值（ standardized value ）符號爲" Z "，也就可寫成 $=\dfrac{x-\mu}{\theta}$。此變數的變換方法就叫做 變換（ Z－transformafion ），變數Z的頻數分佈

稱爲Z分佈（　－distribution ），這種常態分佈曲線就叫做標準常態分佈曲線（ standard normal curve ）。

現舉幾個例子來說明：

〔例1〕設男性成人身高的頻數分佈呈常態分佈，平均身高 $\mu=68$ 吋，標準偏差 $\theta=2$ 吋，求出身高不足70吋的人數佔多少比例？

解：

$$\text{代公式 } Z=\frac{x-\mu}{\theta}\text{，得}$$

$$=\frac{70-68}{2}$$

$$=1$$

查常態分佈機率表，知 $Z=1$ 時，曲線下　到中點的面積爲0.3413，因此，小於70吋的數值是 $0.3413+0.5=0.8413$，也就是佔了84.13％。

〔例2〕試求身高在66吋到70吋之間者佔多大比例？

解：

因爲分佈是呈對稱形，所以身高66吋到68吋的人數等於68吋到70吋的人數。

$(0.5-0.1587)\times 2=0.6826$

68.26％即是身高66吋～70吋的人數

由於許多數據大的分佈均爲常態分佈，所以，在處理數據時就便利許多，可應用標準常態分佈曲線面積表來研究，如身高、體重、脈搏振動都是這一類的例子，但是如人的死亡分佈年齡、家庭的經濟狀況，這些資料都不屬於常態分佈。

另有一很重要的一點，即是全體樣本中平均值很近似常態分佈。

由樣本平均值所構成的全體的平均等於原來全體的平均，（ $\mu\,\overline{x}=\mu$ ），標準偏差等於原來全體的標準偏差除以 $\sqrt{n}$，用符號來表示則

爲 $\theta\overline{x}=\frac{\theta}{\sqrt{n}}$，式中的 n 代表樣本的大小。若 n 合理的變大，如此，它的樣本平均值也就越接近常態分佈。常態分佈的情形是依照其分佈的平均及標準偏差來決定，所以若知道原來全體的 μ 及 θ 之後，樣本平均值的分佈特性也就可得知了。

n 之值應爲多少才算大得合理？可依二種情況來決定：

①我們所希望的樣本平均值分佈接近常態分佈的程度。

②原來分佈的形狀是怎麼樣的？

若其分佈與常態分佈出入很大，樣本數應比原來全體接近常態分佈。

μ 值的大小只決定曲線在 x 軸上的位置，θ 值是決定曲線的形狀，所以爲了比較 θ 值不同的二個常態曲線，便演繹出一種特殊的標準常態曲線（ standard normal curve ）。

在某特定條件下，常態分佈是很接近二項分佈與 Poisson氏分佈的的。如在二項分佈中，不管 π 值爲何，若另一參數 n 增至無限大時，它的分佈情形便接近常態分佈的分佈情形。

如果 $\pi=\frac{1}{2}$，則二項分佈便呈對稱性，所以，π 接近於 $\frac{1}{2}$ 時，二項分佈的形狀要比 π 接近 0 或 1 時更接近常態分佈。所以，如果 n 夠大時，二項變數 r 的分佈可視爲近似常態分佈，這時，平均爲 $n\pi$，而標準偏差爲：

$$\sqrt{\{n\pi(1-\pi)\}}$$

若 Poisson 分佈時，其平均爲 μ，μ 值若增至無限大時，它的形狀也近似常態分佈，所以 Poisson 氏變數 x 的分佈也可視爲接近常態分佈，這時它的平均是 μ，標準偏差是 $\sqrt{\mu}$。

可以採用連續性校正法（ continuity correction ）來運用常態分佈機率來估計二項分佈或 Poisson 氏分佈的近似機率。因爲前者是連續型變數的分佈，後者是間斷型變數的分佈，因此，二項變數 r 的正確機率近似於 $r-\frac{1}{2}$ 至 $r+\frac{1}{2}$ 的常態變數之機率。所以，若二項分佈中

變數大於或等於 r 時，機率便接近標準化常態偏量。

$$= \frac{|r - n\pi| - \frac{1}{2}}{\sqrt{\{n\pi(1-\pi)\}}}$$

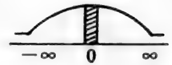
下二表分別列出二項分佈與 Poisson 分佈時的機率運用連續性校正後與常態分佈的比較：

		平均	標準偏差			連續性校正後之常態近似值	
π	n	$n\pi$	$\sqrt{n\pi(1-\pi)}$	r 值	正確機率	z	機率
0.05	10	5	1.581	$\leqslant 2$	0.0537	1.581	0.0579
				$\geqslant 8$	0.0547		
0.1	50	5	2.121	$\leqslant 2$	0.1117	1.179	0.1192
				$\geqslant 8$	0.1221		
0.5	40	20	3.162	$\leqslant 14$	0.0403	1.739	0.0410
				$\geqslant 26$	0.0403		
0.2	100	20	4.000	$\leqslant 14$	0.0804	1.375	0.0846
				$\geqslant 26$	0.0875		

連續性校正後二項分佈與常態分佈的比較

平均 μ	標準偏差 $\sqrt{\mu}$	x 值	正確機率	連續性校正後之常態近似值 $z=\frac{\lvert x-\mu\rvert-\frac{1}{2}}{\sqrt{\mu}}$	機率
5	2.236	0	0.0067	2.013	0.0221
		$\leqslant$ 2	0.1246	1.118	0.1318
		$\geqslant$ 8	0.1334	1.118	0.1318
		$\geqslant$ 10	0.0318	2.013	0.0221
20	4.472	$\leqslant$ 10	0.0108	2.124	0.0168
		$\leqslant$ 15	0.1565	1.006	0.1572
		$\geqslant$ 25	0.1568	1.006	0.1572
		$\geqslant$ 30	0.0218	2.124	0.0168
100	10.000	$\leqslant$ 80	0.0226	1.950	0.0256
		$\leqslant$ 90	0.1714	0.950	0.1711
		$\geqslant$ 110	0.1706	0.950	0.1711
		$\geqslant$ 120	0.0282	1.950	0.0256

連續性校正後，波以松氏分布與常態分布之比較



表（7－2）　常態分配機率表

$W=\frac{X}{\sigma}=$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1271	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2643	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4172	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706

表（7－2） 常態分配機率表（續）

$W=\frac{X}{\sigma}=$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4954	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4980	0.4980	0.4981
2.9	0.4981	0.4982	0.4983	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990									
3.2	0.4993									
3.3	0.4995									
3.4	0.4997									
3.5	0.4998									
3.6	0.49984									
3.7	0.49989									
3.8	0.49993									
3.9	0.49995									
4.0	0.49997									
5.0	0.4999997									

二、機率的演算及意義

機率的意義

首先，先解釋何謂隨機序列（ random sequence 或random series ）？以抛錢幣爲例，拿一枚錢幣抛相當多的次數，而每一次都要記下正或反面，那麼，可能會獲得這樣的結果：

反反正反正正正反正反反正正反………

像這樣的一組合便是隨機序列，序列中的正或反各爲一次抛幣的所得結果，我們稱之爲事件（ event 或outcome ）。隨機序列的特點是根本無一定的排列軌跡可尋，無法預測下一試驗的結果，也可說是每次抛幣的結果都不受前一次或前幾次抛幣所得結果的影響，但它們出現的機率應爲完全相等的，若試驗的次數愈多，那麼某一事件發生的比例的變動愈來愈小而接近於某一界限值，最後的比例就是該事件的機率。

機率的定義，即是某事件發生的機率爲在相同條件下，經過重複的試驗，所得最終的相對頻數（ The probability of an event is the event's longrun relative frequency in repeated trials under similar conditions ）。依據此定義來解釋機率有兩個須特別注意的問題：一爲在相同條件下，經過重複試驗；二爲相對頻數，因此機率的數值是在 0 及 1 之間，所以它的表示法可用小數、分數或者是百分比數。

也可以另一代表符號來表示，如一事件發生的機率爲 P（ A ），則不發生A事件的機率也就是互補事件（ complementary event ）爲 1 － P（ A ）。

機率 Probability，也就是或然率的數學理論應用在醫學上爲統計推論（ statistical inferences ）及統計決策（ statistical

decision ）的工具。

機率的演算

①條件機率及乘法定理 Condfional Probability and the Multipliplicative Law

有時候，某事件的發生會影響到另一事件的發生機率，即以A發生爲條件，B再發生時，這二個事件就是相依事件（conditional events），例如一把牌，它下一張的出現率要以已出現的牌爲依據來決定了。此種相依的條件關係可表示爲P（B｜A），口述爲若事件A已發生時，B事件的發生機率。比如在玩撲克牌時，抽出的第一張牌是紅桃的機率是P（A）＝12/52，若第一張果眞是紅桃，則第二張是紅桃的機率是 $P(B \mid A)=\frac{(13-1)}{(52-1)}=\frac{12}{51}$。

假設A與B爲相依事件，則A與B同時發生的機率爲該兩事件機率的相乘積，這就是所謂的乘法定理（Multiplicative Law），也就是

$$P(A與B)=P(B \mid A)\times P(A)$$

再以上例來說，在一副牌中，連續抽出二張都是紅桃的牌的機率爲：

$$P(A與B)=(\frac{12}{51})(\frac{13}{52})=\frac{156}{2652}=0.0588$$

②互斥事件與加法定理 Mutually Exclusive Events and the Additive Law

互斥是指兩事件不可能同時發生，以前述的拋錢幣爲例，拋出一枚錢幣，若是正面就不可能是反面，若是反面就不可能是正面，絕無兩面同時出現的可能。

若A及B爲互斥事件，則在一次試驗中，發生A或B的機率是該二事件之和，這就是加法定理 Additive Law，以符號表示爲：

$$P(A或B)=P(A)+P(B)$$

若互斥事件是二個以上時，加法定理同樣可成立，可以下式表示之：

P（A或B或C或……）＝P（A）＋P（B）＋P（C）＋……

③獨立事件 Independent Events

獨立事件是指，一事件A的發生與另一事件B的發生完全無關時，這二事件即為獨立事件，以符號表示為：

P（B | A）＝P（B）

例如拋錢幣，不論已試驗多少次，每一次新的試驗，其正反兩面出現的機率恆為 1/2 。若A與B為獨立事件，A及B同時發生的機率可用乘法定理表示：

P（A與B）＝P（A）・P（B）

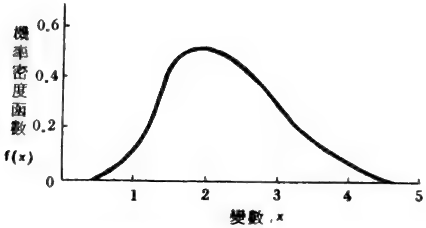
機率分佈 Probability Distributions

例如在一個班級中，男生與女生的組成型態就是一種機率分佈。在機率分佈中的各個不同數值，一般通稱為隨機變數（random variable）。

①連續資料 Continuous data

如果隨機變數是連續型，因為它的變數的機率值一般為0，所以用機率來表示它的變數值是沒有用的。

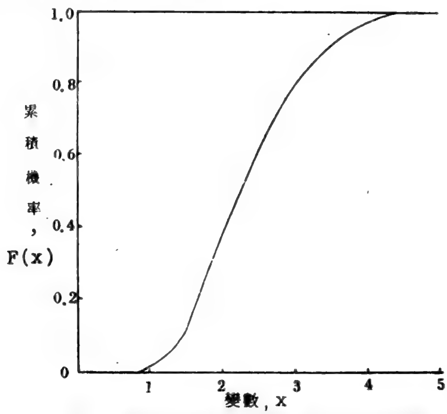
一連續隨機變數x的機率分佈可用如下圖形表示：



連續隨機變數的機率密度函數

應用數學的理論來表示，f（x）是機率F（x）的微分，而隨機變數的數值是小於或等於x，在上圖中，f（x）叫做分佈函數（distribution function），圖中曲線下的面積就是它的數值，自分佈圖的左端起（可以是$-\infty$）到x爲止。

分佈函數與密度函數的關係如下圖：



同上圖中密度函數的分佈函數

密度函數的高度和分佈函數的斜率成正比，於變數值爲中間值時，便到最高點。

機率分佈的形狀可用次數分佈的特性來說明，與衆數的例子一樣，機率或機率密度達最高值，另外還有偏斜的機率分佈。

期望值 Expectation

若機率分佈的觀察值爲不定數（infinite），則該如何來計算其平均值？

若　 目當大時，一個間斷型隨機變數的相對頻數可以機率的近似值表示。

一般若x爲間斷型隨機變數，其個別數值x_0，x_1，x_2，……分別有機率P_0，P_1，P_2………則x的平均值爲：

$$\sum_i x_i p_i$$

應用此法所求得的隨機變數平均值，一般稱之爲期望值（ expected valve，mathematical expectation ），或爲x的期望值（ expectation ），演算式（各x與機率相乘），以符號表示則爲E（ x ），這種隨機變數的期望值常用希臘字母μ（讀音 mu ）來表示，用來區別與固定觀察值的平均值。

若x爲連續性分佈，就不能以E（ x ）＝（出現數A × 0.25 ）＋（出現數B × 0.50 ）＋（出現數C × 0.25 ）計算，但可考慮在間斷型分佈中，每一個x的數值，都有極小的間隔2h，那麼，任何一個x_0都可有一機率，也就是應用連續分佈，x值介於$x_0 - h$與$x_0 + h$之間，在h相當小的時候，幾乎等於2hf（ x ），所以間斷型分佈的x期望值可用普通的方式計算，但在h值極小時，間斷型分佈就更近似連續型分佈，期望值則近於

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx$$

這個公式是一個連續型分佈中x的期望值。

另外，隨機變數的變異數（ variance ）的定義爲：

$$E(x-\mu)^2$$

也就是各數值和平方值差的平方，它的意義和公式$\Sigma(xi - \overline{x})^2/n$一樣。固定觀察數n的變異數，由於此數值是各數值和樣本平均$\overline{x}$差的平方的平均，這時候分母是n或n－1都無所謂了，因爲在n爲無限大時，隨機變數的變異數，一般都以符號σ^2表示，標準偏差的定義就是變異數的開方根σ。

樣本變異數公式的分析：

$$\sigma^2 = E(x)^2 - \mu^2$$

証明：

$$\sigma^2 = E(x-\mu)^2$$
$$= E(x^2 - 2x\mu + \mu^2)$$

$$= E(x)^2 - 2\mu^2 + \mu^2$$

$$= E(x^2) - \mu^2$$

二項分佈The Binomial Distribution

在一隨機序列中，若每一試驗的所得結果互為互斥，即A或B，而以π及$1-\pi$分別表示其發生的機率（π在此不代表圓周率3.14159，而是希臘字母，讀音Pi）。假設在這種隨機序列中，觀察數為n的各種結果。將各類的結果視為n次觀察中的樣本，在樣本中A的出現次數機率是多少？設此數為r，r則為0，1，2，……，n－1，n。同時設定在樣本中出現A的比率（Proportion）$P = r/n$，B出現的比率為：

$$q = (n-r)/n = 1-P$$

r個A及n－r個B出現的機率則為

$$\pi^r(1-\pi)^{n-r}$$

乘以自n個數中取出的r數所得為A之次數，此一係數即稱為二項係數（binomial coefficient）。若n及r的數值頗大時，演算上就較麻煩，一般二項式係數的代表符號是：

$\binom{n}{r}$或${}^nC_r$（讀為n二項式r）計算公式為：

$$\binom{n}{r} = \frac{n(n-1)(n-2)\cdots\cdots(n-r+1)}{1\cdot2\cdot3\cdots\cdots r}$$

上列公式中的分母$1\cdot2\cdot3\cdots\cdots r$稱為r階乘或階乘r（r factorial or factorial r）一般的寫法是r!。而分子$n(n-1)(n-2)\cdots\cdots(n-r+1)$則可改為$\frac{n!}{(n-r)!}$故$\binom{n}{r} = \frac{n!}{r!(n-r)!}$

依二項式的定義，二項式係數為對稱性，故

$$\binom{n}{r} = \binom{n}{n-r}$$

也就是說，從n個物件中取出r個，等於是自n個中取出剩餘的n－r個。

若r＝0，上列公式則無意義，

若r＝n，則

$$\binom{n}{n}=\frac{n!}{n!}=1$$

因對稱的結果，便得：

$$\binom{n}{0}=1$$

由於自n個物件中，取出0物件代表A的方法只有一種，由全體中取出n代表B的方法也只有一種，所以上式的結果是正確的，我們便可確知0！＝1。

前例中的二項係數

$$\binom{4}{0}=1$$

$$\binom{4}{1}=\frac{4}{1}=4$$

$$\binom{4}{2}=\frac{4\cdot 3}{1\cdot 2}=6$$

$$\binom{4}{3}=\frac{4\cdot 3\cdot 2}{1\cdot 2\cdot 3}=4$$

$$\binom{4}{4}=\frac{4\cdot 3\cdot 2\cdot 1}{1\cdot 2\cdot 3\cdot 4}=1$$

使用巴斯卡氏圖形（Pascal's triangle），n若是少數，二項式係數便很容易獲知：

n	1
1	1　1
2	1　2　1
3	1　3　3　1

4	1　4　6　4　1
5	1　5　10　10　5　1
等等	等等

在這三角圖形中，每一數都是上行最近的左右二數之和，可以依此往下推。故每一行中的二項式係數為：

$$\binom{n}{0}，\binom{n}{1}，\cdots\cdots\binom{n}{n-1}，\binom{n}{n}$$

在幾個數的樣本中，具有r個A與n－r個B的機率是：

$$\binom{n}{r}\pi^{r}(1-\pi)^{n-r}$$

假設r為0，1，2，……，n，則上式所得的n＋1的數值總和就表示出現0個A，1個A，2個A……n個A的機率，因為這種的出現為互斥事件，因此，所有的機率總和是1。由理論來計算也得：

$$\binom{n}{0}\pi^{0}(1-\pi)^{n}+\binom{n}{1}\pi^{1}(1-\pi)^{n-1}+$$
$$\cdots\cdots+\binom{n}{n}\pi^{n}(1-\pi)^{0}$$
$$=[\pi+(1-\pi)]^{n}$$
$$=1^{n}$$
$$=1$$

以上兩式的結果完全相同。

r的期望值及變異數

$$E(r)=n\pi$$
$$\mathrm{var}(r)=n\pi(1-\pi)$$

即出現A的平均數等於觀察數乘以A的出現機率。

再討論變異數，假設在n個數中，當$n=1-\pi=\frac{1}{2}$，var(r)達最高值，也就是$\mathrm{var}(r)=\frac{1}{4}n$，又當$\pi$值近於0或1時，它的下降情形頗為明顯，如果$\pi$值很小，因$1-\pi$很接近1，所以var(r)接近$n\pi$，也就是E(r)的數值。

我們在討論樣本中出現A的比率時，也就與P的機率分佈有關。因爲P＝r（1/n），又因1/n的數值在各個樣本中均爲一定數，所以：

$$E(P)=E(r)\times(1/n)=\pi$$

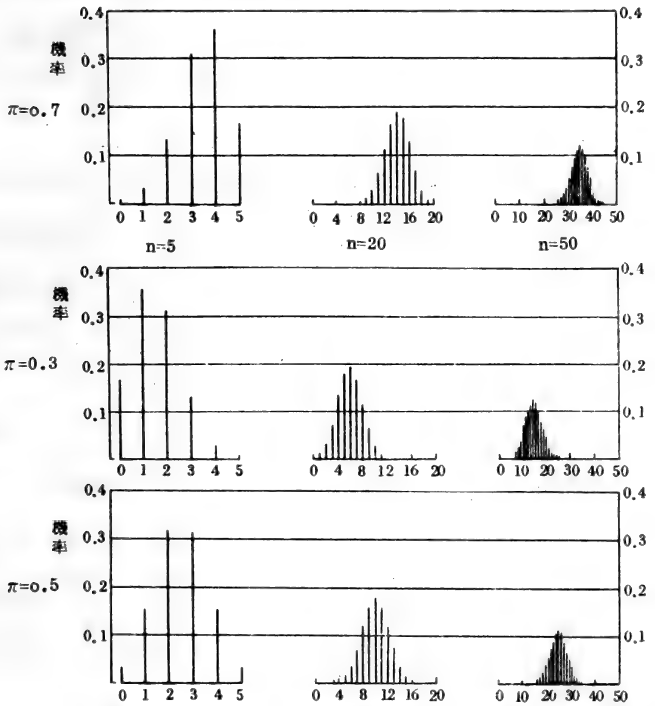
又

$$\mathrm{var}(P)=\mathrm{var}(r)\times(1/n)^{n}$$
$$=\frac{\pi(1-\pi)}{n}$$

也可以r或P的標準偏差來表示：

$$SD(r)=\sqrt{[n\pi(1-\pi)]}\text{，}SD(P)=\sqrt{\left[\frac{\pi(1-\pi)}{n}\right]}$$

二項式分佈的圖形決定於π與n此二參數（parameters）。π與r爲不同的二項式的分佈圖形如下：



π與r不同時的二項分佈，圖中各橫軸表r值

由上圖可知在一特定的n值時，如果$\pi=\frac{1}{2}$，圖形會呈對稱，如果$\pi<\frac{1}{2}$或$>\frac{1}{2}$，圖形就不爲對稱，又在某一特定π值時，n增加時，則不對稱情形會減少。

波以松氏分布The Poisson Distribution

這是爲紀念法國數學家S. D. Poisson 氏（1781～1840）所以命名爲波以松氏分佈，是在研討小樣本的隨機分佈。

設在一隨機狀態下，一事件以平均速率λ的時間重複發生，再假設它發生的時間相當短，以h表示（如千分秒），那麼，某事件發生的機率大祗和h成正比，也就是λh。若在這段時間內，有一種以上的事件發生，它的機率於時間短時可省略不計。

假設在短時間內所發生的事件爲獨立事件，與第一次所發生過的不再重複。現有一很好的例子來說明這類的問題，放射性物質的放射作用，放射性粒子的放射作用λ爲一常數，且粒子的放射爲隨機狀態，這種情形就和獨立事件的隨機序列相同，稱爲Poisson氏過程（Poisson process）。

若在Poisson氏過程中，一事件的發生率爲λ，經重複觀察，延長時間到T，在此段間隔中，以x表示事件發生的次數，不同的時間必有不同的發生次數，也就是說發生的次數爲0，1，2……屬隨機變數的情形。在某一特定x時的機率──→自然是λT，就是以發生的比例乘以時間，λT是x分佈中的平均，可以符號μ表之。

若將時閣T分爲n個段落，使它成爲$\frac{T}{n}$，如果n值很大時，那麼在每個段落的時間內，事件發生的次數可視爲03，也許也可能是1，但大於1就不大可能了。這種情形大抵與二項式n次試驗的排列一樣（在一小段落時間內觀察一次），發生事件爲一次的機率是$\lambda(T/n)=\mu/n$，不發生事件的機率是$1-\mu/n$，經過幾次試驗，在整個序列中，發生x次事件的機率，可用下式表示：

$$\frac{n(n-1)\cdots\cdots(n-x+1)}{x!}\left(\frac{\mu}{n}\right)^{x}\left(1-\frac{\mu}{n}\right)^{n-x}$$

當n增大時，這二項式的近似值就愈近似，所以，若n變爲無限大時，上式中的$n(n-1)\cdots\cdots(n-x+1)$即可以$n^{x}$表示，因爲x比n小就可省略不計。同理，$(1-\mu/n)^{n-x}$，可用$(1-\mu/n)^{n}$表示，因$(1-\mu/n)^{x}$在n增大時，就近於1。就數學上的理論言，若n增至無限大時，$(1-\mu/n)^{n}$就接近e^{-}，e在此爲自然對數(或爲Napierian氏對數)之底($e=2.718\cdots\cdots$)。所以，當n增加到無限大時，發生x事件的機率就近於：

$$P_x=\frac{n^{x}}{x!}\left(\frac{n}{x}\right)^{x}e^{-}=\frac{\mu^{x}e^{-}}{x!}$$

上式是Poisson氏機率分佈的定義，隨機變數x爲0，1，2，⋯⋯，其相對應的機率爲：

$$P_0=e^{-}$$

$$P_1=\mu e^{-}$$

$$P_2=\frac{1}{2}\mu^{2}e^{-}$$

其餘依此類推。

與二項式分佈一樣，若$x=0$，$x!$就等於1，下式爲証明機率總和爲1：

$$\begin{aligned}&P_0+P_1+P_2+\cdots\cdots\\&=e^{-}\left(1+\mu+\frac{1}{2}\mu^{2}+\cdots\cdots\right)\\&=e^{-}\times e\\&=1\end{aligned}$$

由公式$P_x=\dfrac{n^{x}}{x!}\left(\dfrac{n}{x}\right)^{x}e^{-}=\dfrac{\mu^{x}e^{-}}{x!}$中可知機率完全由一參數$\mu$所支配，也就是說$\mu$決定所有的分佈的特性，所以平均值與變異數乃係$\mu$之函數，平均值爲：

$$E(x)=\sum_{x=0}^{\infty} xP_x$$

$$=\sum_{x=0}^{\infty} \frac{x\mu^x e^-}{x!}$$

$$=\sum_{x=1}^{\infty} \frac{\mu^x e^-}{(x-1)!}$$

$$\left[\because \frac{x}{x!}=\frac{1}{(x-1)!}\right]$$

當 $x=0$，總和亦為0

$$E(x)=\mu \sum_{x=1}^{\infty} \frac{\mu^{x-1} e^-}{(x-1)!}$$

$$=\mu \sum_{i=1}^{\infty} \frac{\mu^i e^-}{i!}$$

以 $i=x-1$ 代入，即 $=\mu$

由於總和是包括Poisson氏分佈的所有機率，所以等於1。

變異數的簡單計算式是 $E(x^2)-\mu^2$，

因：

$$E[x(x-1)]=E(x^2)-E(x)$$

$$E[x(x-1)]=\sum_{x=0}^{\infty} x(x-1)P_x$$

$$=\mu^2 \sum_{x=2}^{\infty} \frac{\mu^{x-2} e^-}{(x-2)!}$$

$$=\mu^2$$

同理，

$$E(x^2)-E(x)=\mu^2$$

$$E(x^2)=\mu^2+E(x)$$

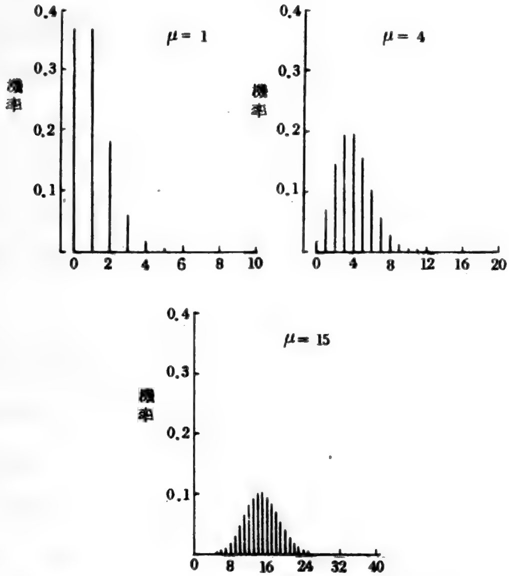
$$=\mu^2+\mu$$

而，

$$\begin{aligned} \mathrm{var}(x) &= E(x^2) - \mu^2 \\ &= (\mu^2 + \mu) - \mu^2 \\ &= \mu \end{aligned}$$

因此，我們可知x的變異數與平均數都等於μ，標準偏差爲$\sqrt{\mu}$。

二項式分佈中，若n增大而π減小，則這個分佈便近於Poisson氏分佈，平均是$\mu = n\pi$，標準偏差在二項分布爲$\sqrt{[n\pi(1-\pi)]}$，在Poisson氏分佈爲$\sqrt{n\pi}$，若π值很小，則標準偏差幾乎相等。



不同μ值時的以松氏分佈，圖中各橫軸表示x值

上圖中顯示$\mu = 1$，4，15的分佈情形，當$\mu = 1$時，它的分佈形狀是十分偏歪的，當$\mu = 4$時，偏斜的情形就少些，當$\mu = 15$時，大概就沒有偏斜的現象了。

第八章
相關係數及迴歸方程式

一、複相關中包括二個問題

複相關

三個或三個以上的變量的相關就是所謂的複相關。天地之間包羅萬象，各事各物之彼此關係又錯綜複雜，息息相關，相互牽引。

若不考慮第三因素，只分析二種變量之間的關係就叫做簡單相關或混合相關。

複相關的意義是必須含括二個問題：一是探討二個或二個以事的變量同時與第三變量之間的關係。另一是假設其他因素相等或不變，並捨棄其他的有關因素，不予硏討，只討論二變量間單純的關係，稱爲純相關或分析相關。因此可知，廣義的複相關也就含括了純相關。複相關可分爲估計標準誤、淨相關係數、多元直線迴歸方程與複相關係數。

①多元直線迴歸

ⓐ迴歸方程式

1 三元

$$x_1' = a + b_{12\cdot3}\,x_2 + b_{13\cdot2}\,x_3$$

公式〔8－1〕

x_1'：應變數 x_1 的估計值

x_2 、 x_3 ：自變數

a：常數項

$b_{12.3}$ ：排除 x_3 的影響後，x_1 對 x_2 的淨迴歸係數

$b_{13.2}$ ：排除 x_2 的影響後，x_1 對 x_3 的淨迴歸係數

$$x_1' = b_{12.3}x_2 + b_{13.2}x_3 = B_{12.3}\frac{S_1}{S_2}x_2 + B_{13.2}\frac{S_1}{S_3}x_3$$

公式〔8－2〕

即：

$$(x_1' - \overline{x}_1) = b_{12.3}(x_2 - \overline{x}_2) + b_{13.2}(x_3 - \overline{x}_3)$$

$$x_2' = b_{21.3}x_1 + b_{23.1}x_3 = B_{21.3}\frac{S_2}{S_1}x_1 + B_{23.1}\frac{S_2}{S_3}x_3$$

$$x_3' = b_{31.2}x_1 + b_{32.1}x_2 = B_{31.2}\frac{S_3}{S_1}x_1 + B_{32.1}\frac{S_3}{S_2}x_2$$

$b_{12.3}$ ：排除 x_3 的影響後，

x_2 ：自變數

x_1 ：附變數的淨迴歸係數

其餘 $b_{13.2}$ 等依此類推

若完全相關時，$b_{12.3}$ 及 $b_{21.3}$ 相等，若不完全相關，二者不相等。

2.四元

$$x_1' = b_{12.34}x_2 + b_{13.24}x_3 + b_{14.23}x_4$$

公式〔8－3〕

3.多元

$$x_1' = b_{12.34\cdots\cdots n}x_2 + b_{13.24\cdots\cdots n}x_3 \cdots\cdots + b_{1n.23\cdots\cdots(n-1)}x_n$$

公式〔8－4〕

ⓑ由相關係數求淨迴歸法

1.三元

$$\triangle = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix}$$

$$\triangle_{11} = \begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix}$$

$$\triangle_{12} = -\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & 1 \end{vmatrix}$$

$$\triangle_{13} = \begin{vmatrix} r_{21} & 1 \\ r_{31} & r_{32} \end{vmatrix} = -\begin{vmatrix} 1 & r_{21} \\ r_{32} & r_{31} \end{vmatrix}$$

$$r_{12} = r_{21}$$

$$r_{23} = r_{32}$$

$$r_{13} = r_{31}$$

$$b_{12\cdot3} = B_{12\cdot3}\frac{S_1}{S_2} = -\frac{\triangle_{12}}{\triangle_{11}}\frac{S_1}{S_2} = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}\frac{S_1}{S_2}$$

公式〔 8 － 5 〕

$$b_{13\cdot2} = B_{13\cdot2}\frac{S_1}{S_3} = -\frac{\triangle_{13}}{\triangle_{11}}\frac{S_1}{S_3} = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}\frac{S_1}{S_3}$$

2.四元

$$b_{12\cdot34} = B_{12\cdot34}\frac{S_1}{S_2}$$

$$= \frac{r_{12} + r_{14}r_{23}r_{34} + r_{13}r_{24}r_{34} - r_{12}r_{34}^2 - r_{13}r_{23} - r_{14}r_{24}}{1 + 2r_{23}r_{24}r_{34} - r_{23}^2 - r_{24}^2 - r_{34}^2}\frac{S_1}{S_2}$$

公式〔 8 － 6 〕

$$b_{13\cdot24} = B_{13\cdot24}\frac{S_1}{S_3}$$

$$= \frac{r_{13} + r_{14}r_{23}r_{24} + r_{12}r_{24}r_{34} - r_{13}r_{24}^2 - r_{12}r_{23} - r_{14}r_{34}}{1 + 2r_{23}r_{24}r_{34} - r_{23}^2 - r_{24}^2 - r_{34}^2}\frac{S_1}{S_3}$$

$$b_{14\cdot23} = B_{14\cdot23}\frac{S_1}{S_4}$$

$$= \frac{r_{14} + r_{13}r_{23}r_{24} + r_{12}r_{23}r_{34} - r_{14}r_{23}^2 - r_{12}r_{24} - r_{13}r_{34}}{1 + 2r_{23}r_{24}r_{34} - r_{23}^2 - r_{24}^2 - r_{34}^2}\frac{S_1}{S_4}$$

3.多元

$$\triangle=\begin{vmatrix} r_{11} & r_{12}\cdots r_{1n} \\ r_{21} & r_{22}\cdots r_{2n} \\ \cdots & \cdots \\ r_{n1} & r_{n2}\cdots r_{nn} \end{vmatrix};\quad \begin{aligned} b_{12.34\cdots n} &= -\frac{\triangle_{12}}{\triangle_{11}}\frac{S_1}{S_2} \\ b_{13.24\cdots n} &= -\frac{\triangle_{13}}{\triangle_{11}}\frac{S_1}{S_3} \\ &\cdots\cdots \\ b_{1n.23\cdots(n-1)} &= -\frac{\triangle_{1n}}{\triangle_{11}}\frac{S_1}{S_n} \end{aligned}$$

公式〔8－7〕

補充說明：

b右下方所標的符號，第一標號是因變數的標號，第二標號是自變數的標號，這二個標號叫做主標號。主標號下以分點隔開的各標號，都是次標號。主標號的次序不能任意倒換，如 $b_{12.34\cdots n}$ 是以 x_1 爲因變數，x_2 是自變數，$b_{21.34\cdots n}$ 是以 x_2 爲因變數，x_1 爲自變數。b的次標號次序則無關重要。有 n′ 個次標號的廻歸係數，叫做 n′ 級廻歸，例如 b_{12} 爲零級廻歸，$b_{12.3}$ 爲一級廻歸，$b_{12.34}$ 爲二級廻歸。

ⓒ用低級廻歸係數求高級淨廻歸係數法

1.三元

$$b_{12.3}=\frac{b_{12}-b_{13}\,b_{32}}{1-b_{23}\,b_{32}}$$

公式〔8－8〕

2.四元

$$b_{12.34}=\frac{b_{12.3}-b_{14.3}\,b_{42.3}}{1-b_{24.3}\,b_{42.3}}$$

公式〔8－9〕

3.多元

$$b_{12.34\cdots n}=\frac{b_{12.34\cdots(n-1)}-b_{1n.34\cdots(n-1)}\,b_{n2.34\cdots(n-1)}}{1-b_{2n.34\cdots(n-1)}\,b_{n2.34\cdots(n-1)}}$$

公式〔8－10〕

由上三式可知，任何級廻歸係數，都可自零級廻歸一級級計算出。

②估計標準誤

簡單相關內的二變量爲 x 及 Y，Y 的估計標準誤

$$S_{y.x}=\sqrt{\frac{\Sigma(Y-Y')^2}{N}}=\sqrt{\frac{\Sigma y'^2}{N}}=S_y\sqrt{1-r^2}$$

本段用 x_1 代替 Y，x_2 代 x，x_1 代 Y′，S_1 代 S_y，$S_{1.2}$ 代 $S_{y.x}$。

S_1：零級標準差

$S_{1.2}$：一級標準誤

如此，$S_{y.x}$ 的公式便可演化爲：

$$S_{1.2}=\sqrt{\frac{\Sigma(x_1-x_1')^2}{N}}$$

$$=\sqrt{\frac{\Sigma(x_1-x_1')^2}{N}}$$

$x_1=x_1-\overline{x}_1$，$x_1'=x_1'-\overline{x}_1$

簡相關估計值 x_1' 是按迴歸直線公式：$x_1'=b_{12}x_2$ 計算出，三元以上的估計標準誤，只要用估計值 x_1' 依公式〔8－2〕，〔8－3〕或〔8－4〕求出，亦可應用下式估計標準誤：

$$S_{1.23\ldots n}=\sqrt{\frac{\Sigma(x_1-x_1')^2}{N}}$$

公式〔8－11〕

公式〔8－11〕中，$S_{1.23\ldots n}$ 是 x_1 對 x_2，x_3…與 x_n 的估計標準誤。若依公式〔8－11〕求多元複相關的估計標準誤，過程太麻煩，可改用淨迴歸係數與其他方法計算估計標準誤比較方便。使用淨迴歸係數計算估計標準誤的公式：

$$S_{1.23}^2=\frac{1}{N}(\Sigma x_1^2-b_{12.3}\Sigma x_1x_2-b_{13.2}\Sigma x_1x_3)$$

公式〔8－12〕

$$S_{1.234}^2=\frac{1}{N}(\Sigma x_1^2-b_{12.34}\Sigma x_1x_2-b_{13.24}\Sigma x_1x_3-b_{14.23}\Sigma x_1x_4)$$

公式〔8－13〕

$$S^2_{1\cdot 23\cdots n}=\frac{1}{N}(\ \Sigma x_1^2-b_{12\cdot 34\cdots n}\ \Sigma x_1 x_2-bx_1 x_2 - b_{13\cdot 24\cdots n}\Sigma x_1 x_3\cdots - b_{1n\cdots 23(n-1)}\Sigma x_1 x_n\ )$$

公式〔 8 － 14 〕

估計標準誤的值一定介於零和零級標準差之間，也可說是估計標準誤一定大於零（或等於零），必小於標準差（或等於標準差），也就是高級的估計標準誤一定小於或等於低級的估計標準誤，以符號表示，則爲：

$$0\leq S^2_{1\cdot 23\cdots n}\leq S^2_{1\cdot 23\cdots(n-1)}\leq\cdots\leq S^2_{1\cdot 23}\leq S^2_{1\cdot 2}\leq S^2_1$$

③淨相關係數

若二變數都和第三變數或第三、第四等變數有相互關係，那麼，這二個變數之間一定受其他因素的影響。控制各因素中 n － 2 個變數，使它固定不變，僅求二個變數間的相關係數，便是淨相關係數或分析相關係數。

ⓐ由低級相關係數求高一級相關係數

$$r_{12\cdot 3}=\frac{r_{12}-r_{13}\,r_{23}}{\sqrt{1-r_{13}^2}\ \sqrt{1-r_{23}^2}}$$

公式〔 8 － 15 〕

$$r_{12\cdot 34}=\frac{r_{12\cdot 3}-r_{14\cdot 3}\,r_{24\cdot 3}}{\sqrt{1-r_{14\cdot 3}^2}\ \sqrt{1-r_{24\cdot 3}^2}}$$

公式〔 8 － 16 〕

$$r_{12\cdot 34\cdots n}=\frac{r_{12\cdot 34\cdots(n-1)}-r_{1n\cdot 34\cdots(n-1)}\,r_{2n\cdot 34\cdots(n-1)}}{\sqrt{1-r_{1n\cdot 34\cdots(n-1)}^2}\ \sqrt{1-r_{2n\cdot 34\cdots(n-1)}^2}}$$

公式〔 8 － 17 〕

擁有 n′ 個次標號的純相關係數，叫做 n′ 級相關係數，如 r_{12} 是零級相關係數。$r_{12\cdot 3}$ 是一級相關係數，$r_{12\cdot 34}$ 是二級相關係數。自上述各公式可得知，n′ 級相關係數可由（ n′ － 1 ）級相關係數求出，也就是由零

級相關係數一級級的推算，便能計算出各高級相關係數。

ⓑ自淨廻歸係數計算淨相關係數

$$r_{12\cdot3}=\sqrt{b_{12\cdot3}\,b_{21\cdot3}}$$

公式〔8－18〕

$$r_{12\cdot34\cdots n}=\sqrt{b_{12\cdot34\cdots n}\,b_{21\cdot34\cdots n}}$$

公式〔8－19〕

ⓒ自估計標準誤計算淨相關係數

$$S^2_{1\cdot23}=S^2_{1\cdot2}(1-r^2_{13\cdot2})=S^2_1(1-r^2_{12})(1-r^2_{13\cdot2})$$

公式〔8－20〕

$$S^2_{1\cdot234}=S^2_1(1-r^2_{12})(1-r^2_{13\cdot2})(1-r^2_{14\cdot23})$$

公式〔8－21〕

$$S^2_{1\cdot23\cdots n}=S^2_{1\cdot23\cdots(n-1)}(1-r^2_{1n\cdot23\cdots(n-1)})$$

$$=S^2_1(1-r^2_{12})(1-r^2_{13\cdot2})\cdots(1-r^2_{1n\cdot23\cdots(n-1)})$$

公式〔8－22〕

$$r^2_{13\cdot2}=1-\frac{S^2_{1\cdot23}}{S^2_{1\cdot2}}$$

公式〔8－23〕

$$S^2_{1\cdot23}=S^2_1(1-r^2_{12})(1-r^2_{13\cdot2})$$

公式〔8－24〕

$$r_{1n\cdot23\cdots(n-1)}=1-\frac{S^2_{1\cdot23\cdots n}}{S^2_{1\cdot23\cdots(n-1)}}$$

公式〔8－25〕

$$S^2_{1\cdot234\cdots n}=S^2_1(1-r^2_{12})(1-r^2_{13\cdot2})(1-r^2_{14\cdot23})\cdots$$

$$(1-r^2_{1n\cdot23\cdots(n-1)})$$

公式〔8－26〕

ⓓ自淨相關係數與估計標準誤計算淨廻歸係數

$$b_{12\cdot3}=r_{12\cdot3}\frac{S_{1\cdot23}}{S_{2\cdot13}}$$

$$b_{12\cdot34} = r_{12\cdot34}\frac{S_{1\cdot234}}{S_{2\cdot134}}$$

公式〔 8 － 27 〕

$$b_{12\cdot34\cdots n} = r_{12\cdot34\cdots n}\frac{S_{1\cdot234\cdots n}}{S_{2\cdot134\cdots n}}$$

公式〔 8 － 28 〕

公式〔 8 － 29 〕

兩變數淨相關係數的範圍介於0 與± 1 之間，正負號表示淨相關的方向，絕對值的大小代表淨相關程度的大小。至於淨相關係數的個數，若k 種變數間不是完全相關，可有kC_2 個淨相關係數。

④複相關係數

複相關係數是一應變數的觀察值及其估計值的簡相關，是一變數與其他二個或二個以上變數的相關係數，一般是以R 表示。依簡相關係數公式是：

$$r_{12} = \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2 \Sigma x_2^2}}$$

$$= \frac{\Sigma x_1 x_1'}{\sqrt{\Sigma x_1^2 \Sigma x_1^2}}$$

當中x' 是自$b_{12}x_2$ 計算出的x_1 之估計值。複相關係數公式可依其爲：

$$R_{1\cdot23\cdots n} = \frac{\Sigma x_1 x_1'}{\sqrt{\Sigma x_1^2 \Sigma x_1'^2}}$$

公式〔 8 － 30 〕

$R_{1\cdot23\cdots n}$是代表應變數x_1對其他自變數x_2 ，x_3 ，$x_4 \cdots x_n$ 之間的複相關係數。當中x_1'是依公式〔 8 － 4 〕計算所得，亦可說是x_1' 是x_2 ，x_3 ，$\cdots x_n$ 等變數一起對x_1的複合估計值。

ⓐ以簡單相關係數求複相關係數

$$R_{1\cdot23} = \sqrt{1-〔(1-r_{12}^2)(1-r_{13\cdot2}^2)}$$

公式〔8－31〕

$$R_{1.23}=\sqrt{1-[(1-r_{13}^{2})(1-r_{12.3}^{2})]}$$

公式〔8－32〕

$$R_{1.23}=\sqrt{1-\frac{\triangle}{\triangle_{11}}}$$

$$=\sqrt{\frac{r_{12}(r_{12}-r_{31}r_{23})+r_{13}(r_{31}-r_{21}r_{32})}{1-r_{23}^{2}}}$$

公式〔8－33〕

$$R_{1.23}=\sqrt{\frac{r_{12}^{2}+r_{13}^{2}-2r_{12}r_{13}r_{23}}{1-r_{23}^{2}}}$$

公式〔8－34〕

$$R_{1.234}=\sqrt{1-[(1-r_{12}^{2})(1-r_{13.2}^{2})(1-r_{14.23}^{2})}$$

公式〔8－35〕

$$R_{1.234}=\sqrt{1-[(1-r_{13}^{2})(1-r_{14.3}^{2})(1-r_{12.34}^{2})]}$$

公式〔8－36〕

$$R_{1.234}=\sqrt{1-[(1-r_{14}^{2})(1-r_{12.4}^{2})(1-r_{12.34}^{2})]}$$

公式〔8－37〕

$$R_{1.234}=\sqrt{1-[(1-r_{12}^{2})(1-r_{14.2}^{2})(1-r_{13.24}^{2})]}$$

公式〔8－38〕

$$R_{1.234}=\sqrt{1-[(1-r_{13}^{2})(1-r_{12.3}^{2})(1-r_{14.23}^{2})]}$$

公式〔8－39〕

$$R_{1.23\cdots n}=\sqrt{1-[(1-r_{12}^{2})(1-r_{13.2}^{2})(1-r_{14.23}^{2})\cdots(1-r_{1n.23\cdots n-1}^{2})]}$$

公式〔8－40〕

$$R_{1.234}=\sqrt{1-[(1-r_{14}^{2})(1-r_{13.4}^{2})(1-r_{12.34}^{2})]}$$

公式〔8－41〕

$$R_{1\cdot23\cdots n}=\sqrt{1-\frac{\triangle}{\triangle_{11}}}$$

公式〔8－42〕

$$\triangle=\begin{vmatrix} r_{11} & r_{12}\cdots\cdots r_{1n} \\ r_{21} & r_{23}\cdots\cdots r_{2n} \\ \cdots\cdots & \cdots\cdots \\ r_{n1} & r_{n2}\cdots\cdots r_{nn} \end{vmatrix}$$

$$=r_{11}\triangle_{11}+r_{12}\triangle_{12}+r_{13}\triangle_{13}+r_{14}\triangle_{14}+\cdots\cdots+r_{1n}\triangle_{1n}$$

ⓑ由淨廻歸係數求複相關係數

$$R_{1\cdot23}=\sqrt{\frac{b_{1\cdot23}\Sigma x_1x_2+b_{13\cdot2}\Sigma x_1x_3}{\Sigma x_1^2}}$$

公式〔8－43〕

$$R_{1\cdot23\cdots n}=\sqrt{\frac{b_{12\cdot34\cdots n}\Sigma x_1x_2+b_{13\cdot24\cdots n}\Sigma x_1x_3\cdots}{\Sigma x_1^2}}$$
$$\overline{+b_{1n\cdot23\cdots(n-1)})\Sigma x_1x_n}$$

公式〔8－44〕

ⓒ以估計標準誤求複相關係數

$$S^2_{1\cdot23}=S_1^2(1-R^2_{1\cdot23})$$

公式〔8－45〕

$$S^2_{1\cdot234}=S_1^2(1-R^2_{1\cdot234})$$

公式〔8－46〕

$$S^2_{1\cdot23\cdots n}=S_1^2(1-R^2_{1\cdot23\cdots n})$$

公式〔8－47〕

$$R^2_{1\cdot23}=1-\frac{S^2_{1\cdot23}}{S_1^2}$$

公式〔8－48〕

$$R^2_{1\cdot 234} = 1 - \frac{S^2_{1\cdot 234}}{S^2_1}$$

公式〔8－49〕

$$R^2_{1\cdot 23\cdots n} = 1 - \frac{S^2_{1\cdot 23\cdots n}}{S^2_1}$$

公式〔8－50〕

ⓓ以B與簡單相關係數求複相關係數

$$R^2_{1\cdot 23} = B^2_{12\cdot 3} + B^2_{13\cdot 2} + 2r_{23}B_{12\cdot 3}B_{13\cdot 2}$$

公式〔8－51〕

$$R^2_{1\cdot 234} = B^2_{12\cdot 34} + B^2_{13\cdot 24} + B^2_{14\cdot 23} + 2r_{23}B_{12\cdot 34} \times B_{13\cdot 24} + 2r_{24}B_{12\cdot 34}B_{14\cdot 23} + 2r_{34}B_{13\cdot 24}B_{14\cdot 23}$$

公式〔8－52〕

$$R^2_{1\cdot 23\cdots n} = r_{12}B_{12\cdot 34\cdots n} + r_{13}B_{13\cdot 24\cdots n} + \cdots\cdots + r_{1n}B_{1n\cdot 23\cdots(n-1)}$$

公式〔8－53〕

ⓔ以淨相關係數求複相關係數

$$1 - R^2_{1\cdot 23} = (1 - r^2_{12})(1 - r^2_{13\cdot 2})$$

公式〔8－54〕

$$1 - R^2_{1\cdot 23\cdots n} = (1 - r^2_{12})(1 - r^2_{13\cdot 2})\cdots(1 - r^2_{1n\cdot 23\cdots(n-1)})$$

公式〔8－55〕

複相關係數的大小是代表自變數對應變數影響程度的大小，並不能代表相關的正負，所以複相關係數介於0與1之間（$0 \leq R \leq 1$）。

另一組變數和同一應變數的高級複相關係數一定大於或等於它的低級複相關係數（$0 \leq R^2_{1\cdot 2} \leq R^2_{1\cdot 23} \cdots \leq R^2_{1\cdot 23\cdots n}$）；而且，同一組變數與同一應變數的複相關係數一定大於或等於淨相關係數的絕對值，若 $R^2_{1\cdot 23} \geq r^2_{13\cdot 2}$，當 $r_{12} = 0$，亦即 x_1 及 x_2 無關係時，$R^2_{1\cdot 23} = r^2_{13\cdot 2}$。

複相關係數的個數，是k種變數的複相關，在變數間的不完全相關

時，可計算出k 個複相關係數。

二、二變量資料間之數值的增長具有互相一致者謂之正相關

二種變量之間的關係，是呈直線的，叫做直線相關（ Linear correlation ），是呈曲線的，就叫做曲線相關或非直線相關（ Nonlinear correlation ）。二變量之間的數值一樣的，叫做正相關（ Positive correlation ），不一樣的，叫做負相關（ Negative correlation ）。再者，一變量數值的增減，另一變量數值增減不定，方向不定的，稱為無相關或是零相關（ No Correlation ）。

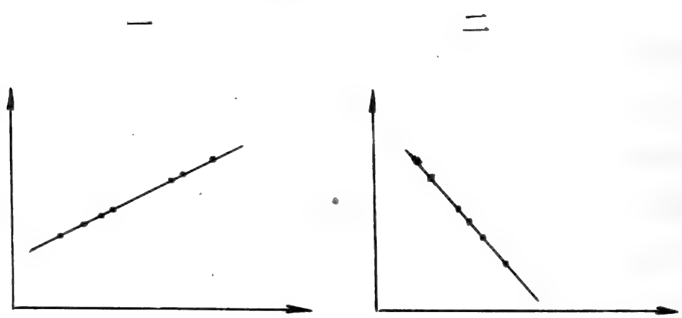
①簡單相關

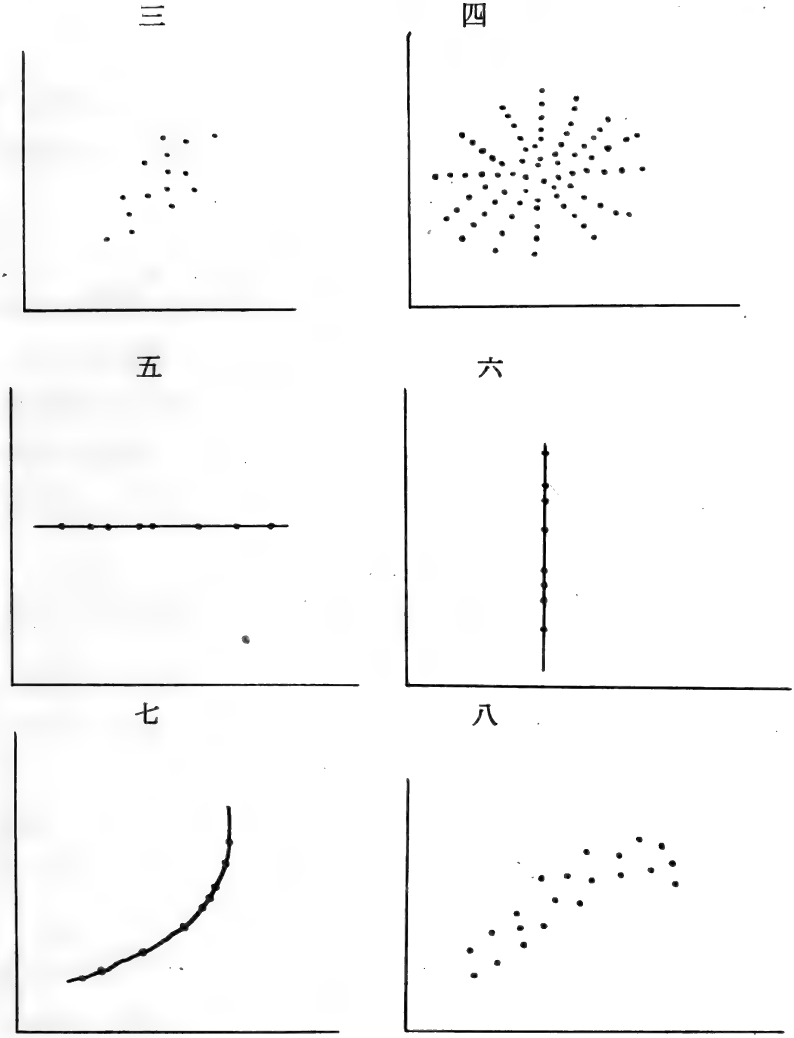
ⓐ散布圖

由原變量看二資料間的關係，不容易看出二者的相關情形，若繪製散布圖（ Scatter Diagram ），就比較容易一目了然了。繪製散布圖，是用橫軸及縱軸各代表一變量，於圖中點出各對數值的位置。

二變量之間是否有關係？相關的方向是正是負？相關程度是高是低？以及相關的是直線相關或曲線相關？由散布圖中就可一窺全貌了。

下圖（ 8 － 1 ）是各種相關的散布情形，可資參考。





圖（8－1）

一圖的點排列於右上角至左下角的直線上，屬於直線完全正相關。

二圖的點排列於右下角至左上角的直線上，屬於直線完全負相關。

三圖的點散布在直線兩邊，分布極廣，是屬中度直線相關。

四圖是呈放射狀的分布，五圖各點排列在一橫線上，六圖各點排列

在一縱線上，都是無相關。

七圖的各點排列在一條曲線上，是屬於完全曲線相關。

八圖的各點位於曲線兩邊的附近位置，是中度以上的曲線相關。圖中顯示的各點愈集中在斜直線或曲線附近的位置，相關的程度就愈高，分布的位置範圍愈大，相關程度就愈小。

②直線的積差相關

英國統計學家Karl Pearson 計算相關係數是依二種變量各與其平均數的差數乘積的平均數來求出，所以也叫做積差相關係數（ The Product－Moment Coefficient of Correlation ）。這種積差相關係數爲學者普遍採用，一般是用r表示。r的數值一定介於0與±1之間，相關係數的絕對值越大，相關程度也就愈高。通常由0至±0之間將r分爲幾個等級，表示不同的相關程度。

若N是5對以下，$r = 0.8$，也許並不顯著，若對數是在200對以上，$r = 0.2$，則顯著。

1. r　0 —— ±0.2　無相關
2. r ±0.2 —— ±0.4　低相關
3. r ±0.4 —— ±0.6　中度相關
4. r ±0.6 —— ±0.8　顯著相關
5. r ±0.8 —— ±1　高相關

上述的分級，只做爲鑑定相關度高低的參考，不是絕對的。一般在社會現象或生物現象方面，二變量之積差相關很少能達±0.7以上。

ⓐ計算未分組資料的積差相關係數

未分組的資料，要計算積差相關，可應用下列公式來求出：

$$r = \frac{\Sigma xy}{N S x S y}$$

公式〔8－56〕

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\Sigma y^2}}$$

公式〔8－57〕

r：是表示積差相關係數

N：項數

S_x：X變量的標準差　$S_x=\sqrt{\frac{\Sigma(X-M_x)^2}{N}}=\sqrt{\frac{\Sigma x^2}{N}}$

S_y：Y變量的標準差　$S_y=\sqrt{\frac{\Sigma(Y-My)^2}{N}}=\sqrt{\frac{\Sigma y^2}{N}}$

x：X變量及其平均的差數　$x=X-M^x$

y：Y變量及其平均的差數　$y=Y-My$

公式〔8－56〕是積差相關係數的基本公式，由於積差相關係數的定義，原來是二種變數及其平均數的差的積數的平均數，即$\frac{\Sigma xy}{N}$，但這樣會受二變量單位的影響，致使$\frac{\Sigma xy}{N}$的數值變動不定，爲了彌補此項缺點，可以其標準差爲計算單位。

公式〔8－57〕是演化公式〔8－56〕而得，可省略求標準差的程序，較爲方便，因此計算未分組資料的積差相關時，應多利用公式〔8－57〕。

若二變量的平均數是整數，使用公式〔8－57〕是較爲方便，但一般二變量的平均數都不是整數，所得到的差數也一樣，所以在計算上非常麻煩，在此情況下，可使用下列公式：

$$r=\frac{N\Sigma XY-(\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2-(\Sigma x)^2][N\Sigma Y^2-(\Sigma Y)^2]}}$$

公式〔8－58〕

$dx=X-Ax$，

$dy=Y-Ay$，

即：

$X = dx + Ax$

$Y = dy + Ay$

（ Ax與Ay均爲假定平均數）

ⓑ計算已分組資料的積差相關係數

如果二變量資料的相對項數很多，而不加以分組歸類而使用未分組資料計算相關係數的公式，很浪費時間，且容易發生錯誤，所以必須先將二資料分類整理交叉分組表（ Bivariable frequenoy distribution ），然後再計算，這要方便許多。這種計算積差相關係數的相關表（ Correlation table ）的編製方法，一是二種變量成對數值分組，一是單一變量數值分組。

相關表的製作過程：

①算出二變量的全距，各用二變量的最大值，減去最小值。

②各確定二變量的組距與組限。

③製作相關表的畫記表，X變量橫排在表的上方，組值小的在左，大的在右。Y變量縱排在表的左方，組值小者在下，大者在上。

④以畫記法將二變量成對數值分別畫入適當的方格內。

⑤計算每一方格內的次數，其次將同行及同列方格次數相加，得到各行各列的次數，最後算出各行各列的總次數，必須要相等於原變量成對數值總項數才無誤，若不相等，就要重新檢查，直到結果相等爲止。

$$r = \frac{\Sigma\Sigma f'(X-M_x)(Y-M_y)}{N S_x S_y}$$

公式〔8－59〕

$$r = \frac{\Sigma\Sigma f'_{xy}}{\sqrt{(\Sigma f_x X^2)(\Sigma f_y Y^2)}}$$

公式〔8－60〕

$$r=\frac{N\Sigma\Sigma f'XY-(\Sigma f_xX)(\Sigma f_yY)}{\sqrt{[N\Sigma f_xX^2-(\Sigma f_xX)^2][N\Sigma f_yY^2-(\Sigma f_yY)^2]}}$$

公式〔8－61〕

S_x：X變量的標準差

S_y：Y變量的標準差

f'：各小方格內的次數

f_x：X變量各組的次數

f_y：Y變量各組的次數

X與Y：變量的原始數值

M_x：X變量的平均數

M_y：Y變量的平均數

$x=X-M_x$

$y=Y-M_y$

公式〔8－59〕是計算分組資料積差相關的基本公式。公式〔8－60〕與公式〔8－61〕都是演化公式〔8－59〕而來的，公式〔8－60〕是由差數計算相關，公式〔8－61〕是以原變量計算相關。

這三個公式的計算都很麻煩，求分組資料積差相關都不應用這三個公式，而以假定平均數計算積差相關，較爲方便，公式如下：

$$r=\frac{N\Sigma\Sigma f'd_xd_y-(\Sigma f_xd_x)(\Sigma f_yd_y)}{\sqrt{[N\Sigma f_xd_x^2-(\Sigma f_xd_x)^2][N\Sigma f_yd_y^2-(\Sigma d_yf_y)^2]}}$$

公式〔8－62〕

$d_x=X-A_x$

$d_y=Y-A_y$

A_x是X變量的假定平均數，A_y是Y變量的假定平均數。若X變量的各組組距相同，Y變量的各組組距相同，將公式〔8－62〕改爲下式〔8－63〕，計算上較方便：

$$\gamma=\frac{N\Sigma d'_y\ \Sigma f'\ d'_x-(\Sigma f_x d'_x)(\Sigma f_y d'_y)}{\sqrt{[N\Sigma f_x d'^2_x-(\Sigma f_x d'_x)^2][N\Sigma f_y d'^2_y-(\Sigma f_y d'_y)^2]}}$$

公式〔8－63〕

如果二變量的標準差已知，以組距為單位，便可使用下列公式計算積差相關：

$$\gamma=\frac{\dfrac{\Sigma d'_y\ \Sigma f'\ d'_x}{N}-\dfrac{\Sigma f_x d'_x}{N}\ \dfrac{\Sigma f_y d'_y}{N}}{S'_x\ S'_y}$$

$$=\frac{\dfrac{\Sigma d'_y\ \Sigma f'\ d'_x}{N}-\dfrac{\Sigma f_x d'_x}{N}\ \dfrac{\Sigma f_y d'_y}{N}}{\sqrt{\dfrac{\Sigma f_x d'^2_x}{N}-\left(\dfrac{\Sigma f_x d'_x}{N}\right)^2}\ \sqrt{\dfrac{\Sigma f_y d'^2_y}{N}-\left(\dfrac{\Sigma f_y d'_y}{N}\right)^2}}$$

公式〔8－64〕

S'_x：X 變量的標準差，以組距為單位者

S'_y：Y 變量的標準差，以組距為單位者

積差相關係數的準確度，一般都算至二位小數，也有算至三位小數的。

ⓒ積差相關及直線迴歸

由迴歸係數 $b_{yx}=\dfrac{\Sigma xy}{\Sigma x^2}$ 與 $b_{xy}=\dfrac{\Sigma xy}{\Sigma y^2}$ 可知道，計算迴歸方程式，原本不需要相關係數，但迴歸係數 b_{yx} 與 b_{xy} 和積差相關係數 γ 關係密切。因為 $b_{yx}=\dfrac{\Sigma xy}{\Sigma x^2}$，$b_{xy}=\dfrac{\Sigma xy}{\Sigma y^2}$，$\gamma=\dfrac{\Sigma xy}{NS_xS_y}$，因此，可化解成 $b_{yx}=\gamma\dfrac{S_y}{S_x}$，$b_{xy}=\gamma\dfrac{S_x}{S_y}$，所以迴歸方程式的公式可改成：

$$Y'=\gamma\frac{S_y}{S_x}(X-M_x)+M_y$$

公式〔8－65〕

公式〔8－55〕的迴歸方程式公式可改成

$$X'=\gamma\frac{S_x}{S_y}(Y-M_y)+M_x$$

公式〔8－66〕

迴歸方程式具有某種程度的預測功能，這種預測的結果雖然不一定和觀察值完全吻合，但至少在目前是最可靠的一種方法。

ⓓ估計標準誤

由X值可估計Y值Y′，但估計得到的Y′的值和觀察所得的Y值間常有出入，這種差數平均和的平均數之平方根叫做Y的估計標準誤（Standard error of estimate），一般以 S_{yx} 來表示。X的估計標準誤以 S_{xy} 來表示。因此，估計標準誤的計算方法和標準差的計算方式相同，前者是以估計值為中心，後者是以平均數為中心。

$$S_{yx}=\sqrt{\frac{\Sigma(Y-Y')^2}{N}}$$

$$=\sqrt{\frac{\Sigma y'^2}{N}}$$

公式〔8－67〕

$$S_{xy}=\sqrt{\frac{\Sigma(X-X')^2}{N}}$$

$$=\sqrt{\frac{\Sigma x'^2}{N}}$$

公式〔8－68〕

估計標準誤也可使用下列公式：

$$S_{yx}=S_y\sqrt{1-r^2}$$

公式〔8－69〕

$$S_{xy}=S_x\sqrt{1-r^2}$$

公式〔8－70〕

觀察公式〔8－69〕與公式〔8－70〕，若 $\gamma=0$，估計標準誤

和標準差相等，這時，估計標準誤的值最大，換句話說是估計值的正確性最低。由此可知相關愈小，估計標準誤愈和標準差相近，迴歸方程式的估計價值愈小。若 $\gamma = \pm 1$，估計標準誤的最小數值是0，這就表示各觀察值和估計值一樣，都是在迴歸線上，沒有估計誤差，估計值的正確性最高。由此可知相關愈大，估計標準誤的數值愈接近0，迴歸方程式的估計價值愈大。

公式〔8－69〕、〔8－70〕中的 $\sqrt{1 - r^2}$，一般是用k來表示，叫做不相關係數，最大值是1，最小值是0，可以表示不相關程度的大小。1－k是預測效力指數，可表示迴歸方程式預測效果的大小。

若二變量是直線相關，並且成對項數多時，各估計誤差所成的次數分配，近似常態分配，可依常態分配法則，來推論預測的可靠與否。

預測的可靠性的推算方法，在進行這種可靠性推斷時，必須依大量項數求出的迴歸方程式與估計標準誤。

使用一變量的觀察值及估計值求 γ

$$\gamma = \frac{\Sigma yy'}{\sqrt{\Sigma y^2 \Sigma y'^2}}$$

使用迴歸係數求 γ

$$\gamma = \pm\sqrt{b_{xy} \times b_{yx}}$$

（γ 和 b_{xy} 與 b_{yx} 同號）

使用估計標準誤求 γ

$$\gamma = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}}$$

$$= \sqrt{1 - \frac{S_{xy}^2}{S_x^2}}$$

③等級相關

如果統計資料只有等級或質量上的大小次序時，可使用等級相關法

來求相關。再者，如果變量項數不多（30以下），使用此法亦極爲簡便。等級相關雖然不比積差相關來得正確，但却方便許多。此法是爲司畢門Spearman所提出，分爲等級計差法（Method of rankdifferences）與等級計餘法（Spearman's food rule）二種方法，其中以等級計差法的正確度較高，公式如下：

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

公式〔8－71〕

ρ：等級計差相關係數的符號，讀音爲Rho

D：X與Y二變量的等級差數

計算等級相關，應注意下列二點：

①若原數值是連續變量化爲等級時，遇有相同數值，應將其所占等級平均計算。

②等級之差數D可省略標注符號，D^2後不受影響。

ρ值及γ值往往極爲相近，求出ρ後，γ就不一定要求出，但爲了使人易於瞭解或使用上的方便起見，可將ρ化成γ。ρ化爲γ的公式如下：

$$\gamma = 2\sin\left(\frac{\pi}{6}\rho\right)$$

公式〔8－72〕

$\pi = 180^\circ$

$\frac{\pi}{6} = 30^\circ$

下表爲正弦、餘弦的函數表：

正角弦度	正弦 0′	10′	20′	30′	40′	50′	60′	餘角弦度
	60′	50′	40′	30′	20′	10′	餘弦 0′	
0	0.0000	0.0029	0.0058	0.0087	0.0116	0.0145	0.0175	89
1	.0175	.0204	.0233	.0262	.0291	.0320	.0349	88
2	.0349	.0378	.0407	.0436	.0465	.0494	.0523	87
3	.0523	.0552	.0581	.0610	.0640	.0669	.0698	86
4	.0698	.0727	.0756	.0785	.0814	.0843	.0872	85
5	.0872	.0901	.0929	.0958	.0987	.1016	.1045	84
6	.1045	.1074	.1103	.1132	.1161	.1190	.1219	83
7	.1219	.1248	.1276	.1305	.1334	.1363	.1392	82
8	.1392	.1421	.1449	.1478	.1507	.1536	.1564	81
9	.1564	.1593	.1622	.1650	.1679	.1708	.1736	80
10	.1736	.1765	.1794	.1822	.1851	.1880	.1908	79
11	.1908	.1937	.1965	.1994	.2022	.2051	.2079	78
12	.2079	.2108	.2136	.2164	.2193	.2221	.2250	77
13	.2250	.2278	.2306	.2334	.2363	.2391	.2419	76
14	.2419	.2447	.2476	.2504	.2532	.2560	.2588	75
15	.2588	.2616	.2644	.2672	.2700	.2728	.2756	74
16	.2756	.2784	.2812	.2840	.2868	.2896	.2924	73
17	.2924	.2952	.2979	.3007	.3035	.3062	.3090	72
18	.3090	.3118	.3145	.3173	.3201	.3228	.3256	71
19	.3256	.3283	.3311	.3338	.3365	.3393	.3420	70
20	.3420	.3448	.3475	.3502	.3529	.3557	.3584	69
21	.3584	.3611	.3638	.3665	.3692	.3719	.3746	68
22	.3746	.3773	.3800	.3827	.3854	.3881	.3907	67
23	.3907	.3934	.3961	.3987	.4014	.4041	.4067	66

正角弦度	正弦0′	10′	20′	30′	40′	50′	60′	餘角弦度
	60′	50′	40′	30′	20′	10′	餘弦0′	
24	.4067	.4094	.4120	.4147	.4173	.4200	.4226	65
25	.4226	.4253	.4279	.4305	.4331	.4358	.4384	64
26	.4384	.4410	.4436	.4462	.4488	.4514	.4549	63
27	.4540	.4566	.4592	.4617	.4643	.4669	.4695	62
28	.4695	.4720	.4746	.4772	.4797	.4823	.4848	61
29	.4848	.4874	.4899	.4924	.4950	.4975	.5000	60
30	.5000	.5025	.5050	.5075	.5100	.5125	.5150	59
31	.5150	.5175	.5200	.5225	.5250	.5275	.5299	58
32	.5299	.5324	.5348	.5375	.5398	.5422	.5446	57
33	.5446	.5471	.5495	.5519	.5544	.5568	.5592	56
34	.5592	.5616	.5640	.5664	.5688	.5712	.5736	55
35	.5736	.5760	.5783	.5807	.5831	.5854	.5878	54
36	.5878	.5901	.5925	.5948	.5972	.5995	.6018	53
37	.6018	.6041	.6065	.6088	.6111	.6134	.6157	52
38	.6157	.6180	.6202	.6225	.6248	.6271	.6293	51
39	.6293	.6316	.6338	.6361	.6383	.6406	.6428	50
40	.6428	.6450	.6472	.6494	.6517	.6539	.6561	49
41	.6561	.6583	.6604	.6626	.6648	.6670	.6691	48
42	.6691	.6713	.6734	.6756	.6777	.6799	.6820	47
43	.6820	.6841	.6862	.6884	.6905	.6926	.6947	46
44	.6947	.6967	.6988	.7009	.7030	.7050	.7071	45
45	.7071	.7092	.7112	.7133	.7153	.7173	.7193	44
46	.7193	.7214	.7234	.7254	.7274	.7294	.7314	43
47	.7314	.7333	.7353	.7373	.7392	.7412	.7431	42

正角	正弦0′	10′	20′	30′	40′	50′	60′	餘角
弦度	60′	50′	40′	30′	20′	10′	餘弦0′	弦度
48	.7431	.7451	.7470	.7490	.7509	.7528	.7547	41
49	.7547	.7566	.7585	.7604	.7623	.7642	.7660	40
50	.7660	.7679	.7698	.7716	.7735	.7753	.7771	39
51	.7771	.7790	.7808	.7826	.7844	.7862	.7880	38
52	.7880	.7898	.7916	.7934	.7951	.7969	.7986	37
53	.7986	.8004	.8021	.8039	.8056	.8073	.8090	36
54	.8090	.8107	.8124	.8141	.8158	.8175	.8192	35
55	.8192	.8208	.8225	.8241	.8258	.8274	.8290	34
56	.8290	.8307	.8323	.8339	.8355	.8371	.8337	33
57	.8387	.8403	.8418	.8434	.8450	.8465	.8480	32
58	.8480	.8496	.8511	.8526	.8542	.8557	.8572	31
59	.8572	.8587	.8601	.8615	.8631	.8646	.8660	30
60	.8660	.8675	.8689	.8704	.8718	.8732	.8746	29
61	.8746	.8760	.8774	.8788	.8802	.8816	.8829	28
62	.8829	.8843	.8857	.8870	.8884	.8897	.8910	27
63	.8910	.8923	.8936	.8949	.8962	.8975	.8988	26
64	.8988	.9001	.9013	.9026	.9038	.9051	.9063	25
65	.9063	.9075	.9088	.9100	.9112	.9124	.9135	24
66	.9135	.9147	.9159	.9171	.9182	.9194	.9205	23
67	.9205	.9216	.9228	.9239	.9250	.9261	.9272	22
68	.9272	.9283	.9293	.9304	.9315	.9325	.9336	21
69	.9336	.9346	.9356	.9367	.9377	.9387	.9397	20
70	.9397	.9407	.9417	.9426	.9436	.9446	.9455	19
71	.9455	.9465	.9474	.9483	.9492	.9502	.9511	18

正角弦度	正弦0′	10′	20′	30′	40′	50′	60′	餘角弦度
	60′	50′	40′	30′	20′	10′	餘弦0′	
72	.9511	.9520	.9528	.9537	.9546	.9555	.9563	17
73	.9563	.9572	.9580	.9588	.9596	.9605	.9613	16
74	.9613	.9621	.9628	.9636	.9644	.9652	.9659	15
75	.9659	.9667	.9674	.9681	.9689	.9696	.9703	14
76	.9703	.9710	.9717	.9724	.9730	.9737	.9744	13
77	.9744	.9750	.9757	.9763	.9769	.9775	.9781	12
78	.9781	.9787	.9793	.9799	.9805	.9811	.9816	11
79	.9861	.9822	.9827	.9833	.9838	.9843	.9848	10
80	.9848	.9853	.9858	.9863	.9868	.9872	.9877	9
81	.9877	.9881	.9886	.9890	.9894	.9899	.9903	8
82	.9903	.9907	.9911	.9914	.9918	.9922	.9925	7
83	.9925	.9929	.9932	.9936	.9939	.9942	.9945	6
84	.9945	.9948	.9951	.9954	.9957	.9959	.9962	5
85	.9962	.9964	.9967	.9969	.9971	.9974	.9976	4
86	.9976	.9978	.9980	.9981	.9983	.9985	.9986	3
87	.9986	.9988	.9989	.9990	.9992	.9993	.9994	2
88	.9994	.9995	.9996	.9997	.9997	.9998	.9998	1
89	.9998	.9999	.9999	1.0000	1.0000	1.0000	1.0000	0

下表為ρ值化為γ值表：

ρ	γ	ρ	γ
0.01	0.0105	0.25	0.2611
0.02	0.0209	0.26	0.2714
0.03	0.0314	0.27	0.2818
0.04	0.0419	0.28	0.2922
0.05	0.0524	0.29	0.3025
0.06	0.0628	0.30	0.3129
0.07	0.0733	0.31	0.3232
0.08	0.0838	0.32	0.3335
0.09	0.0942	0.33	0.3439
0.10	0.1047	0.34	0.352
0.11	0.1151	0.35	0.3645
0.12	0.1256	0.36	0.3748
0.13	0.1360	0.37	0.3850
0.14	0.1465	0.38	0.3935
0.15	0.1569	0.39	0.4056
0.16	0.1674	0.40	0.4158
0.17	0.1778	0.41	0.4261
0.18	0.1882	0.42	0.4363
0.19	0.1986	0.43	0.4465
0.20	0.2091	0.44	0.4567
0.21	0.2195	0.45	0.4669
0.22	0.2299	0.46	0.4771
0.23	0.2403	0.47	0.4872
0.24	0.2507	0.48	0.4973

ρ	γ	ρ	γ
0.49	0.5075	0.75	0.7654
0.50	0.5176	0.76	0.7750
0.51	0.5277	0.77	0.7847
0.52	0.5378	0.78	0.7943
0.53	0.5479	0.79	0.8039
0.54	0.5580	0.80	0.8135
0.55	0.5680	0.81	0.8230
0.56	0.5781	0.82	0.8325
0.57	0.5881	0.83	0.8421
0.58	0.5981	0.84	0.8516
0.59	0.6081	0.85	0.8610
0.60	0.6180	0.86	0.8705
0.61	0.6280	0.87	0.8799
0.62	0.6379	0.88	0.8893
0.63	0.6478	0.89	0.8986
0.64	0.6577	0.90	0.9080
0.65	0.6676	0.91	0.9173
0.66	0.6775	0.92	0.9269
0.67	0.6873	0.93	0.9359
0.68	0.6971	0.94	0.9451
0.69	0.7069	0.95	0.9543
0.70	0.7167	0.96	0.9635
0.71	0.7265	0.97	0.9727
0.72	0.7363	0.98	0.9818
0.73	0.7460	0.99	0.9909
0.74	0.7557	1.00	1.0000

④曲線相關

曲線相關又名非直線相關(Non − linear correlation)。二變量是成直線或曲線可由二方面得知，一是由散布圖的情形，另一是比較直線相關係數和曲線相關係數的差數。γ 是直線相關係數的符號，η_{yx} 是曲線相關係數的符號，叫做相關比(Correlation Ratio)讀音是 eta。γ 是Y對X的相關比(The correlation ratio of Y on X)，後者是X對Y的相關比。二變量之間的直線相關，不管那一個是自變數，那一個是因變數，相關係數或爲正，或爲負只有一個，所以 γ 下不需要再加附標區分。至於相關比，僅作爲表示相關程度的大小，並無正負之分。同一事件的相關比常比它的相關係數 γ 大，而 γ 的絕對值永不可能大於它的相關比。若二變量間呈曲線相關，η 與 γ 的差異顯著，若二變量的關係呈直線相關，η 與 γ 的差異則很小。直線相關係數與曲線相關係數之比的絕對值必介於1與0之間，絕不能大於1，1是完全相關，0是完全不相關。

$$\eta_{yx} = \sqrt{1 - \frac{S_{yw}^2}{S_y^2}}$$

公式〔 8 − 73 〕

$$\eta_{yx} = \sqrt{\frac{S^2_{\bar{y}}}{S_y^2}}$$

公式〔 8 − 74 〕

$$\eta_{xy} = \sqrt{1 - \frac{S_{xw}^2}{S_x^2}}$$

公式〔 8 − 75 〕

$$\eta_{xy} = \sqrt{\frac{S^2_{\bar{x}}}{S_x^2}}$$

公式〔 8 − 76 〕

η_{yx} ：Y對X的相關比

η_{xy} ：X對Y的相關比

S_y^2 ：Y變量標準差的平方，也就是Y變量的變異數。

(Variance)

S_x^2 ：X變量的變異數

S_{yw}^2 ：各縱行數值及其平均數之差數平方和的平均數，亦爲各縱行內各數值及其平方數之差數的變異數。

$$S_{yw}^2=\frac{1}{N}\Sigma\Sigma f_i'(Y_i-\overline{Y}_i)^2$$

公式〔8－77〕

f_i' ：第i縱行內各小方格內的次數

Y_i ：Y變量各組的組距中點

$\overline{Y}_i$ ：第i縱行內的各變量的平均數

$$\overline{Y}_i=\frac{1}{f_i}\Sigma f_i' Y_i$$

公式〔8－78〕

f_i ：第i縱行內的次數，亦爲求直線相關時之相關表中的f_x

$S^2_{\overline{y}}$：縱行內各變量的平均數的變異數

$$S^2_{\overline{y}}=\frac{1}{N}\Sigma f_i(\overline{Y}_i-M_y)^2$$

公式〔8－79〕

$$S^2_{\overline{y}}=\frac{\Sigma f_i d^2}{N}-\left(\frac{\Sigma f_i d}{N}\right)^2$$

公式〔8－80〕

$d=\overline{Y}_i-A$

S_w^2 ：各橫行內數值及其平均數之差數的變異數

$$S_{xw}^2=\frac{1}{N}\Sigma\Sigma f_i'(x_i-\overline{x}_i)^2$$

公式〔8－81〕

$$\overline{X}_i = \frac{1}{f_i}\Sigma f'_i X_j$$

公式〔8－82〕

$$S^2_{\overline{x}} = \frac{1}{N}\Sigma f_i(\overline{X}_i - M_x)^2$$

公式〔8－83〕

$$S^2_{\overline{x}} = \frac{\Sigma f_i d^2}{N} - \left(\frac{\Sigma f_i d}{N}\right)^2$$

公式〔8－84〕

計算 η_{xy}（X對Y的相關比）的過程和計算 η_{yx} 的一樣，不予重述。

二變量的相關，若是完全直線相關，$\eta_{yx}=\gamma=1$，爲曲線完全不相關時，$\eta_{yx}=\gamma=0$，爲曲線相關時，$\eta_{yx}>\gamma$，爲不完全直線相關時，η_{yx} 稍大於 γ，所以 γ 的絕對值，恆不可能大於其相關比，可以下列二式來說明：

$$\eta^2_{yx} = 1 - \frac{S^2_{yw}}{S^2_y}$$

$$\gamma^2 = 1 - \frac{S^2_{yx}}{S^2_y}$$

S^2_{yw}：各縱行數值及其實際平均數之差數的變異數

S^2_{yx}：各縱行數值及其預測平均數的差數的變異數

由於各縱行數值及其實際平均數之差數的平方和爲一最小值，因此，η^2_{yx} 便大於 γ。

三、迴歸方程式與其配合原則

偏態係數、峯度係數、差異數、平均數皆爲一單獨變量的分析。若二種或二種以上之間的資料，都相互有關係，可自當中的一個變數來推算另一個變數，若應用方程式來顯示變量和變量間的關係，則爲迴歸（Regression）方程式，有直線迴歸與曲線迴歸二種，其中曲線迴歸較爲複雜，而直線迴歸較簡單。

兩組變量間相互關係的一直線就是迴歸直線，一般在X和Y二變量之中有兩條迴歸直線，一爲X對Y的迴歸線，另一條就是Y對X的迴歸線。依據X對Y的迴歸線可由Y估計　。依據Y對X的迴歸線可由x估計Y，也就是X爲自變數，Y爲應變數。這二條迴歸線，在二變量有絕對的關係時，二者合而爲一。

確定迴歸直線，即迴歸方程式的配合，必須注意下列二項原則：

①估計值和觀察值的差之代數和爲零，也就是估計值的總和與觀察值的總和相等。

②最小平方法（Method of least squares），使得散布圖中的各點與此直線間垂直距離的平方和爲最小，也就是觀察值與估計值的差的平方和最小。

Y對X的迴歸線（Regression line of Y on X）方程式，用原始變量表示則：

$$Y' = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}(X - M_x) + M_y$$

公式〔8 －85〕

X與Y都是原變量。

M_x與M_y是X變量與Y變量的算術平均數（Y'爲Y的估計值）。

用差異數表示則：

$$y' = \frac{\Sigma xy}{\Sigma x^2} x$$

公式〔8－86〕

$$y' = Y' - M_y$$

$$y = Y - M_y$$

$$x = X - M_x$$

公式〔8－85〕中的 $\frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma x^2 - (\Sigma X)^2}$ 與公式〔8－86〕中的 $\frac{\Sigma xy}{\Sigma x^2}$ 叫做Y對X的迴歸係數（Coefficient of regression of Y on X），一般是以 b_{yx} 表示。

X對Y的迴歸方程式，用原始變量表示時則：

$$X' = \frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}(Y - M_y) + M_x$$

公式〔8－87〕

用差數表示則：

$$x' = \frac{\Sigma xy}{\Sigma y^2} y$$

公式〔8－88〕

$\frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}$ 或 $\frac{\Sigma xy}{\Sigma y^2}$ 爲X對Y的迴歸係數，一般以 b_{xy} 表示。

曲線迴歸

二變數之間的關係不會是全都如上述般的單純，有時會成曲線的關係，要用曲線來配合（Curve fitting）求曲線迴歸（Curvilinear regression），便於估計。

$$Y' = a + bX + cX^2$$

$$Na + b\Sigma X + c\Sigma X^2 = \Sigma Y$$

$$a\Sigma X + b\Sigma X^2 + c\Sigma X^3 = \Sigma XY$$

$$a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 = \Sigma X^2 Y$$

公式〔 8 － 89 〕

從實際資料中各X與Y變量，演算以上聯立方程式，得三未知數a，b，c便可算出X對Y的二次迴歸曲線方程式。

第九章
生存表的意義及應用

一、追踪調查

追踪調查 Follow－Up studies

研究運用生命表的方式來追踪調查如何分析病患的生存型態（ survival pattern ）。這種方法原則上可應用在研究最後非歸於死亡的情形。就如疾病在緩和後病徵出現的某類非致死性問題，實際上，發生這類的問題可能較好，例如經過治療後病徵因而消失。

分析患者的追踪調查時，一般須經歷不同的時間，有的已經開刀過好長的一段時間，有的是近期方動過手術，也有的病人已經死亡，所以分析起來是很準確的，有的在分析時仍然活著，有的人在複查時漏失了，不能繼續追踪調查，原因很多，有的也因醫學上的問題而停止追踪調查，也有因為其他的疾病，或是意外的死亡，無法完成追查工作。

若將上述的各種可能性都排除，又假設受追踪調查的每位病患都能接受調查至死亡，那麼，用生命表的方法來說明它經手術後生存時間的問題便很容易。生命表中的生存率 ℓ_x 等於 ℓ_0 乘上大於 x 的生存比率。所以主要工作在於知道生存時間的分佈，為了解決不完全的資料，以下表來說明：

① 追踪年數	② 追踪報告 死亡	③ 退出	④ 年初生存者	⑤ 危險患者	⑥ 死亡機率	⑦ 生存機率	⑧ x 年後生存率
x～x+1	d_x	w_x	n_x	n'_x	q_x	p_x	ℓ_x
0－1	90	0	374	374.0	0.2406	0.7594	100
1－2	76	0	284	284.0	0.2676	0.7324	75.9
2－3	51	0	208	208.0	0.2452	0.7548	55.6
3－4	25	12	157	151.0	0.1656	0.8344	42.0
4－5	20	5	120	117.5	0.1702	0.8298	35.0
5－6	7	9	95	90.5	0.0773	0.9227	29.1
6－7	4	9	79	74.5	0.5537	0.9463	26.8
7－8	1	3	66	64.5	0.0155	0.9845	25.4
8－9	3	5	62	59.5	0.0504	0.9496	25.0
9－10	2	5	54	51.5	0.0388	0.9612	23.7
10－	21	26	47	—	—	—	22.8

表9－1　癌症患者追踪調查，生命表式

①欄，時間間隔的選擇依資料的性質來決定，在研究調查中必須有手術後逐年的生存率，這只分析到十年，但若需要可延長調查時間。時間的間隔並不一定要一樣，爲了避免表格太過繁複，可將間隔定爲二年一組。

②③欄，是將患者上次的情形分別記於各年的結果，若病人已死亡，便列入第②欄，若病患於上回的調查時仍然活著，便列入第③欄，以退出（ Withdrawn ）來表示還活著的病患，不管他是不是還接受複查，或者是因爲其他因素不能繼續調查。

④欄，是在某一年開始時還活著的病患人數，也就是第②、第③兩欄相加的總數。

⑤欄，表示在 x 到 x ＋ 1 年之間危險的病患的調整數，是：

$$n'_x = n_x - \frac{1}{2} W_x$$

⑥欄，表示在 x 到 x ＋ 1 年之間，死亡機率的估計數值：

$$q_x = d_x \,/\, n'_x$$

因爲中止調查的病患 W_x 在這段時間內，有的人可能會隨時死亡，因此要將 n_x 調整爲 n'_x。若 W_x 病患並未中止調查，便可以預估增加 $\frac{1}{2} q_x W_x$ 的死亡，那麼，所有的死亡人數卽 $d_x + \frac{1}{2} q_x W_x$ ，所以，死亡的機率就可爲：

$$q_x = \frac{d_x + \frac{1}{2} q_x W_x}{n_x}$$

這個公式，實際上和上述的二個公式是一樣的。

⑦欄，生存的機率 $p_x = 1 - q_x$

⑧欄，如果估計手術之後三年的生存機率是 $p_0 \, p_1 \, p_2$ ，一般稱最後一欄爲生命表生存率（ life table survival rate ），與第⑦欄數值的連乘積相等，用 $\ell_0 = 100$ ，公式如下：

$$\ell_x = \ell_0 p_0 p_1 \cdots p_{x-1}$$

這計算法有二種假設，一是，假設停止調查的病患的死亡機率和未中止調查的人數一樣。這是合理的假設，因爲中止調查的病患若仍然接受調查那麼死亡的人數是可以知道的，病患若不能做追踪調查可能和他的健康狀況有關係，所以假設也有危險。

二是，p_x 值是病患在各種不同時間內所調查的計算，我們要假設在各種不同的時間內，機率仍是十分穩定的，否則生命表便不合使用了。

若 d_{10} 與 w_{10} 都爲 0 時，表示十年以後已無病患，n_{10} 是0，q_{10} 與 p_{10} 就不能計算，這時 ℓ_{11} 就不能得到，表示超過十年追踪調查已不能知道生存的資料，若還有病人生存者除外。所以，第⑧欄是生存型態的方法，$\ell_x = 50\%$時的 X 是表示中位生存時間。

演算生命表的 p_x 值，也會受抽樣變動的影響，若不是中止調查的問題，它的變動可視爲二項分佈的型態，樣本數是n_x，中止調查的影響大抵和將樣本數減少到n'_x時一樣，這時候 ℓ_x 的變異數如下：

$$\mathrm{var}(\ell_x) = \ell_x^2 \sum_{i=0}^{x-1} \frac{d_i}{n'_i(n'_i - d_i)}$$

Reed－Merrell換算表

$$5q_x = 1 - e^{-5\,{}_5m_x - 008(5)^3\,{}_5m_x^2}$$

$5m_x$	$5q_x$	△	$5m_x$	$5q_x$	△
		.00			.00
.000	.000 000	4 989	.022	.104 599	4 506
.001	.004 989	4 965	.023	.109 105	4 485
.002	.009 954	4 943	.024	.113 590	4 464
.003	.014 897	4 920	.025	.118 054	4 444
.004	.019 817	4 897	.026	.122 498	4 423
.005	.024 714	4 876	.027	.126 921	4 402
.006	.029 590	4 852	.028	.131 323	4 382
.007	.034 442	4 830	.029	.136 705	4 361
.008	.039 272	4 808	.030	.140 066	4 341
.009	.044 080	4 786	.031	.144 407	4 321
.010	.048 866	4 763	.032	.148 728	4 301
.011	.053 629	4 742	.033	.153 029	4 281
.012	.058 371	4 720	.034	.157 310	4 261
.013	.065 091	4 698	.035	.161 571	4 241
.014	.067 789	4 676	.036	.166 812	4 221
.015	.072 465	4 655	.037	.170 033	4 201
.016	.077 120	4 633	.038	.174 234	4 182
.017	.081 753	4 612	.039	.178 416	4 162
.018	.086 365	4 590	.040	.182 578	4 143
.019	.090 955	4 570	.041	.186 721	4 123
.020	.095 525	4 547	.042	.190 844	4 104
.021	.100 072	4 527	.043	.194 948	4 085

$$5q_x = 1 - e^{-5\,{}_5m_x - 008(5)3\,{}_5m_x^2}$$

$5m_x$	$5q_x$	△	$5m_x$	$5q_x$	△
		.00			.00
.044	.199 033	4 066	.067	.287 866	3 647
.045	.203 099	4 047	.068	.291 513	3 630
.046	.207 146	4 028	.069	.295 143	3 613
.047	.211 174	4 008	.070	.298 756	3 596
.048	.215 182	3 990	.071	.302 352	3 579
.049	.219 172	3 972	.072	.303 931	3 562
.050	.223 144	3 952	.073	.309 493	3 545
.051	.227 096	3 935	.074	.313 038	3 523
.052	.231 031	3 915	.075	.316 566	3 511
.053	.234 946	3 897	.076	.320 077	3 495
.054	.238 843	3 879	.077	.323 572	3 478
.055	.242 722	3 861	.078	.327 050	3 461
.056	.246 583	3 842	.079	.330 511	3 446
.057	.250 425	3 824	.080	.333 956	3 429
.058	.254 249	3 807	.081	.337 385	3 412
.059	.258 056	3 788	.082	.340 797	3 396
.060	.261 844	3 770	.083	.344 193	3 380
.061	.265 614	3 753	.084	.347 573	3 364
.062	.269 367	3 735	.085	.350 937	3 347
.063	.273 102	3 717	.086	.354 284	3 332
.064	.276 819	3 700	.087	.357 616	3 316
.065	.280 519	3 682	.088	.360 932	3 300
.066	.284 201	3 665	.089	.364 232	3 284

$${}_5q_x = 1 - e^{-5\,{}_5m_x - .008(5)^3\,{}_5m_x^2}$$

$_5m_x$	$_5q_x$	△	$_5m_x$	$_5q_x$	△
		.00			.00
.090	.367 516	3 268	.113	.438 851	2 926
.091	.370 784	3 253	.114	.441 777	2 911
.092	.374 037	3 237	.115	.444 688	2 897
.093	.377 274	3 222	.116	.447 585	2 883
.094	.380 496	3 206	.117	.450 468	2 870
.095	.383 702	3 191	.118	.453 338	2 855
.096	.383 893	3 176	.119	.456 193	2 842
.097	.390 069	3 160	.120	.459 035	2 827
.098	.393 229	3 145	.121	.461 862	2 815
.099	.396 374	3 130	.122	.464 677	2 800
.100	.399 504	3 116	.123	.467 477	2 787
.101	.402 620	3 100	.124	.470 264	2 773
.102	.405 720	3 087	.125	.473 037	2 760
.103	.408 805	3 070	.126	.475 797	2 746
.104	.411 875	3 056	.127	.478 543	2 733
.105	.414 931	3 041	.128	.481 276	2 720
.106	.417 972	3 026	.129	.483 996	2 707
.107	.420 998	3 011	.130	.486 703	2 693
.108	.424 009	2 998	.131	.489 396	2 680
.109	.427 007	2 982	.132	.492 076	2 667
.110	.429 989	2 969	.133	.494 743	2 655
.111	.432 958	2 953	.134	.497 398	2 641
.112	.435 911	2 940	.135	.500 039	2 628

$$5q_x = 1 - e^{-5\,{}_5m_x - 008(5)^3\,{}_5m_x^2}$$

5m_x	5q_x	△	5m_x	5q_x	△
		.00			.00
.136	.502 667	2 616	.159	.559 692	2 336
.137	.503 283	2 603	.160	.562 028	2 324
.138	.507 886	2 590	.161	.564 352	2 313
.139	.510 476	2 577	.162	.566 665	2 301
.140	.513 053	2 565	.163	.568 965	2 290
.141	.515 618	2 552	.164	.571 256	2 279
.142	.518 170	2 540	.165	.573 535	2 267
.143	.520 710	2 527	.166	.575 802	2 257
.144	.523 237	2 515	.167	.578 059	2 245
.145	.525 752	2 503	.168	.580 304	2 234
.146	.528 255	2 490	.169	.582 538	2 223
.147	.530 745	2 478	.170	.584 761	2 211
.148	.533 223	2 466	.171	.586 972	2 201
.149	.535 689	2 454	.172	.589 173	2 190
.150	.538 143	2 442	.173	.591 363	2 180
.151	.540 585	2 430	.174	.593 543	2 168
.152	.543 015	2 418	.175	.595 711	2 157
.153	.545 433	2 406	.176	.597 868	2 147
.154	.547 839	2 394	.177	.600 015	2 137
.155	.550 233	2 382	.178	.602 152	2 125
.156	.552 615	2 371	.179	.604 277	2 115
.157	.554 986	2 359	.180	.606 392	2 105
.158	.557 345	2 347	.181	.608 497	2 094

$$_5q_x = 1 - e^{-5\,_5m_x - 008(5)^3\,_5m_x^2}$$

$_5m_x$	$_5q_x$	△	$_5m_x$	$_5q_x$	△
		.00			.00
.182	.610 591	2 083	.205	.655 970	1 856
.183	.612 674	2 073	.206	.657 826	1 847
.184	.614 747	2 063	.207	.659 673	1 838
.185	.616 810	2 053	.208	.661 511	1 829
.186	.618 853	2 042	.209	.663 340	1 819
.187	.620 905	2 032	.210	.665 159	1 810
.188	.622 937	2 022	.211	.666 969	1 802
.189	.624 959	2 012	.212	.668 771	1 792
.190	.626 971	2 002	.213	.670 563	1 783
.191	.628 973	1 992	.214	.672 346	1 774
.192	.630 965	1 982	.215	.674 120	1 765
.193	.632 947	1 972	.216	.675 885	1 756
.194	.634 919	1 962	.217	.677 641	1 747
.195	.636 881	1 952	.218	.679 388	1 739
.196	.638 833	1 943	.219	.681 127	1 729
.197	.640 776	1 933	.220	.682 856	1 721
.198	.642 709	1 923	.221	.684 577	1 712
.199	.644 632	1 913	.222	.686 289	1 704
.200	.646 545	1 904	.223	.687 993	1 695
.201	.648 449	1 894	.224	.689 688	1 686
.202	.650 343	1 885	.225	.691 374	1 678
.203	.652 228	1 876	.226	.693 052	1 699
.204	.654 104	1 866	.227	.694 721	1 661

$$5q_x = 1 - e^{-5\,{}_5m_x - .008(5)^3\,{}_5m_x^2}$$

$5m_x$	$5q_x$	△	$5m_x$	$5q_x$	△
		.00			.00
.228	.696 382	1 652	.251	.732 330	1 469
.229	.698 034	1 644	.252	.733 799	1 462
.230	.699 678	1 636	.253	.735 261	1 453
.231	.701 314	1 627	.254	.736 714	1 447
.232	.702 941	1 619	.255	.738 161	1 439
.233	.704 560	1 611	.256	.739 600	1 432
.234	.706 171	1 602	.257	.741 032	1 424
.235	.707 773	1 595	.258	.742 456	1 417
.236	.709 368	1 586	.259	.743 873	1 409
.237	.710 954	1 578	.260	.745 282	1 403
.238	.712 532	1 570	.261	.746 685	1 395
.239	.714 102	1 562	.262	.748 080	1 388
.240	.715 664	1 555	.263	.749 468	1 381
.241	.717 219	1 546	.264	.750 849	1 374
.242	.718 765	1 538	.265	.752 223	1 366
.243	.720 303	1 531	.266	.753 589	1 360
.244	.721 834	1 522	.267	.754 949	1 353
.245	.723 356	1 515	.268	.756 302	1 345
.246	.724 871	1 507	.269	.757 647	1 339
.247	.726 378	1 500	.270	.758 986	1 332
.248	.727 878	1 492	.271	.760 318	1 325
.249	.729 370	1 484	.272	.761 643	1 318
.250	.730 854	1 476	.273	.762 961	1 311

$$_5q_x = 1 - e^{-5\,{}_5m_x - .008(5)^3\,{}_5m_x^2}$$

$_5m_x$	$_5q_x$	△	$_5m_x$	$_5q_x$	△
		.00			.00
.274	.764 272	1 304	.297	.792 621	1 157
.275	.765 576	1 298	.298	.793 778	1 151
.276	.766 874	1 291	.299	.794 929	1 145
.277	.768 165	1 284	.300	.796 014	1 139
.278	.769 449	1 278	.301	.797 213	1 133
.279	.770 727	1 271	.302	.798 346	1 128
.280	.771 998	1 264	.303	.799 474	1 121
.281	.773 262	1 258	.304	.800 595	1 115
.282	.774 520	1 251	.305	.801 710	1 110
.283	.775 771	1 245	.306	.802 820	1 103
.284	.777 016	1 239	.307	.803 923	1 098
.285	.778 255	1 231	.308	.805 021	1 092
.286	.779 486	1 226	.309	.806 113	1 087
.287	.780 712	1 219	.310	.807 200	1 080
.288	.781 931	1 213	.311	.808 280	1 075
.289	.783 144	1 206	.312	.809 355	1 070
.290	.784 350	1 201	.313	.810 425	1 063
.291	.785 551	1 193	.314	.811 488	1 059
.292	.786 744	1 188	.315	.812 547	1 052
.293	.787 932	1 182	.316	.813 599	1 047
.294	.789 114	1 175	.317	.814 646	1 042
.295	.790 289	1 169	.318	.815 688	1 036
.296	.791 458	1 163	.319	.816 724	1 030

$$_5q_x = 1 - e^{-5\,{}_5m_x - 008(5)^3\,{}_5m_x^2}$$

$_5m_x$	$_5q_x$	△	$_5m_x$	$_5q_x$	△
		.00			.000
.320	.817 754	1 026	.343	.840 001	907
.321	.818 780	1 019	.344	.840 918	903
.322	.819 199	1 015	.345	.841 821	897
.323	.820 814	1 009	.346	.842 718	893
.324	.821 823	1 003	.347	.843 611	888
.325	.822 826	0 999	.348	.844 499	884
.326	.823 825	0 993	.349	.845 383	878
.327	.824 818	0 988	.350	.846 261	874
.328	.825 806	0 982	.351	.847 135	869
.329	.826 788	0 978	.352	.848 004	865
.330	.827 766	0 972	.353	.848 869	860
.331	.828 738	0 967	.354	.849 729	856
.332	.829 705	0 962	.355	.850 585	850
.333	.830 667	0 957	.356	.851 435	847
.334	.831 624	0 952	.357	.852 282	842
.335	.832 576	0 947	.358	.853 124	837
.336	.833 523	0 941	.359	.853 961	833
.337	.834 464	0 937	.360	.854 794	828
.338	.835 401	0 932	.361	.855 622	824
.339	.836 333	0 927	.362	.856 446	819
.340	.837 260	0 922	.363	.857 265	816
.341	.838 182	0 917	.364	.858 081	810
.342	.839 099	0 912	.365	.858 891	807

$$5q_x = 1 - e^{-5\,{}_5m_x - 008(5)^3\,{}_5m_x^2}$$

$5m_x$	$5q_x$	△	$5m_x$	$5q_x$	△
		.000			.000
.366	.859 698	802	.389	.877 092	708
.367	.860 500	798	.390	.877 800	705
.368	.861 298	793	.391	.878 505	700
.369	.862 091	789	.392	.879 205	697
.370	.862 880	785	.393	.879 902	693
.371	.863 665	781	.394	.880 595	689
.372	.864 446	776	.395	.881 284	686
.373	.865 222	773	.396	.881 970	682
.374	.865 995	768	.397	.882 652	678
.375	.866 763	764	.398	.883 300	674
.376	.867 567	760	.399	.884 004	671
.377	.868 287	756	.400	.884 675	667
.378	.869 043	751	.401	.885 342	663
.379	.869 794	748	.402	.886 005	660
.380	.870 542	744	.403	.886 665	656
.381	.871 286	739	.404	.887 321	653
.382	.872 025	736	.405	.887 974	649
.383	.872 761	732	.406	.888 623	646
.384	.873 493	728	.407	.889 269	642
.385	.874 221	723	.408	.889 911	638
.386	.874 944	720	.409	.890 549	635
.387	.875 664	716	.410	.891 184	632
.388	.876 380	712	.411	.891 816	628

$$5q_x = 1 - e^{-5\,{}_5m_x - 008(5)^3\,{}_5m_x^2}$$

$5m_x$	$5q_x$	△	$5m_x$	$5q_x$	△
		.000			.000
.412	.892 444	625	.435	.905 978	550
.413	.893 069	621	.436	.906 528	548
.414	.893 690	618	.437	.907 076	544
.415	.894 308	614	.438	.907 620	541
.416	.894 922	612	.439	.908 161	539
.417	.895 534	607	.440	.908 700	535
.418	.896 141	605	.441	.909 235	532
.419	.896 746	601	.442	.909 767	530
.420	.897 347	598	.443	.910 297	526
.421	.897 945	594	.444	.910 823	524
.422	.898 539	592	.445	.911 347	521
.423	.899 131	588	.446	.911 867	518
.424	.899 719	585	.447	.912 386	514
.425	.900 304	581	.448	.912 900	513
.426	.900 885	579	.449	.913 413	509
.427	.901 464	575	.450	.913 922	
.428	.902 039	572			
.429	.902 611	569			
.430	.903 180	566			
.431	.903 746	562			
.432	.904 308	560			
.433	.904 868	556			
.434	.905 424	554			

二、生命表的內容

生命表亦稱爲死亡表(mortality table)，也可稱做生存－死亡表(death – survival table)，有二種型式，一爲世代生命表(generation life table)，另一名稱爲集團生命表(cohort life table)，是調查一特定年間所出生的人數，依每年變化的衞生狀態下，紀錄死亡人數，追踪調查至這年的人口全部死亡爲止，具有眞實性是這類生命表的優點，但所須耗費的時間太長，所以使用價値並不高。

二是當代生命表(current life table)，一般以生命表簡稱之，是假設一個出生世代，世假設這一個出生世代人數遭受到某一時期的眞實人口所遭受的年齡別死亡風險，人口壽命長短的測度乃依按此統計表來完成記載。在假定的人口數同一時期出生，這些人口在一特定年間各年齡別死亡機率(probability of dying)生存抑或死亡，逐年調查，一直到這些人口全部死亡爲止，這二項假定須依據下列二原則：㈠衞生狀態保持不變。㈡這些人口一直居住在同一地區，不他遷。

生命表常被用來評價一個地區或國度裏的衞生狀態，也就是依照該地區或國度的人口壽命評斷其各種急慢性疾病、傳染病、衞生狀況是否有效的爲人管理。

完全生命表(complete life table)是指生命表的內容含括各年齡的數値。此外，若生命表是每五年、十年或更久分成若干組的就稱爲簡易生命表(abridged life table)。簡易生命表準確性較差，但是省時省力，很適合公共衞生上的應用，若需要調查某一特定年齡的人口數値，可使用內插法來獲知。通常簡易表的內容是0到4歲分成各組各不同年齡，5歲以上方依5歲年齡成組別。

①生存率（probability of living，surviving rate，p_x）

生存率 p_x 是指滿x歲的人，再生存至滿x＋1歲的機率，它和死亡率二者的關係可用下式表之：

$$p_x = 1 - q_x$$

②生存數（number of living，ℓ_x）

生存到x歲時仍活著的人數就是所謂的生存數 ℓ_x，生命表中，最初假設的人口一般都設有100,000人出生，也就是 ℓ_0。

生存數通常在最初的一年中減少得最多，那是因爲嬰兒的死亡率較高之故，後幾年因某某原因，人口死亡越來越加增，生存數也就一再減少。生存數減少到原來出生總數的一半時，女性約爲75歲，男性約爲70歲。在假定的100,000人口中，生存至100歲尙活著的人數男性約有20人，女性約有91人。

③死亡機率（probability of dying，rate of mortality，g_x）

某年齡的人口在該年中可能死亡的風險（risks）或機會稱爲死亡率或死亡機率，一般使用的表示符號爲q，在q的右下角加註年齡數。

未滿一歲的嬰兒死亡率相當高，大約與男性60歲的死亡率相等，過了一歲以後，死亡率便很快減少，死亡率最少的年齡大約在10歲左右，之後死亡率又漸漸增加，30歲左右大概就維持平衡狀態，40歲以上死亡率又增加頗速，隨著年齡的增加，死亡率也不斷地提高，大致說來女性的死亡率一直都比男性爲低。

死亡率並不是每個國家的情形都一樣，這與一國的醫藥水準、生活水準；人民體質都有關連，死亡率的變化有顯著的不同。

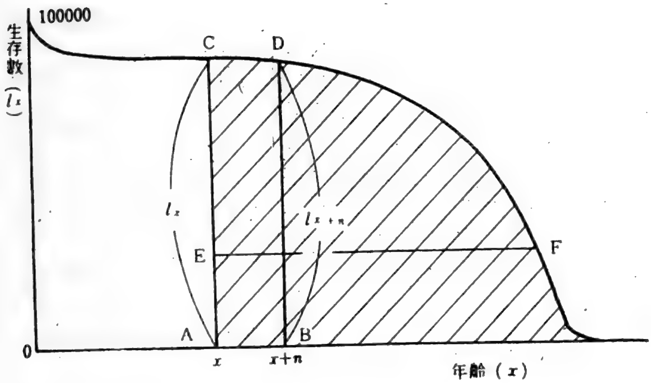
④定常人口（stationary population）

假定一人口集團的出生數爲100,000人，而死亡的秩序不變，經過一段時期，這些人口的年齡組合仍無變化，這種情形就是定常人口，下圖所顯示的是x歲以上、未滿x＋n歲的定常人口，以 ${}_nL_x$ 代表A、B

之間 ℓ_x 曲線下的面積，以 T_x 代表x歲以上的定常人口（斜線部份），以 L_x 代表x歲的定常人口，稱做 $_1L_x$。在AC線上任取一點E，劃一條直線與X軸平行交叉 ℓ_x 曲線於F，EF便表示某人在x歲後的生存年數，也就是可將 T_x 看成 ℓ_x 的人口在x歲後的生存年數總和。以下是 T_x 與 $_nL_x$ 的演算公式：

$$T_x = \int_x^{\infty} \ell_t \, d^t = \sum_{t=x}^{\infty} L_t$$

$$_nL_x = \int_x^{n+x} \ell_t \, d^t$$



⑤平均壽命（exepctation of life, life expectancy, $\mathring{e}_x$）

平均壽命亦稱爲預期壽命，是指假設此時衞生狀態不變，滿x歲尚存活的人口（ℓ_x），以後一直生存到死亡的個人平均生存年數，也就是以後可預期的平均壽命言。

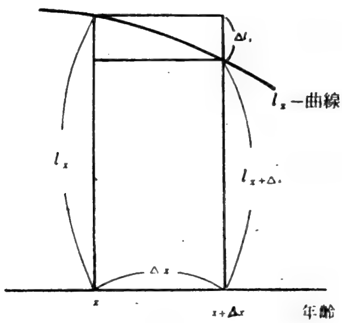
平均壽命自然是隨著年齡的增加而逐漸降低，但未滿一歲的常比一歲的壽命短，其原因已述於前，所以，一歲後的生存者已渡過危險的嬰兒期，壽命便會漸漸提高。一國的知識水準、教育程度、衞生觀念、醫學發展、經濟情況在在都是影響平均壽命的要素。

⑥死力（ force of mortality , μ_x ）

所謂死力，也就是死亡率 ${}_nq_r$ 是 x 歲到 x ＋ n 歲之間的死亡機率，可以下列公式表之：

$${}_nq_r = \frac{\ell_x - \ell_{x+n}}{\ell_x}$$

n 是帶動改變整個數值的原動力，應用這種方式來表示 ℓ_x 的變化情形並不完全，所以，要測定 x 歲的 ℓ_x 減少程度，應該用相對變化率來表示會更恰當些。



如上圖中 x 與 x ±△x （△x ＜0 ）的相對生存數各是 ℓ_x ，則可以下式公式表示 ℓ_x ＋△x 對 X 變化的 ℓ_x 的相對變化率 m ：

$$m = -\frac{1}{\ell_x} \cdot \frac{\ell_x + \triangle x - \ell_x}{\triangle x}$$

$$= -\frac{1}{\ell_x} \cdot \frac{\triangle \ell_x}{\triangle x}$$

$$\triangle x \rightarrow 0 \text{ 則 } \lim_{\triangle x \rightarrow 0} \left(-\frac{1}{\ell_x} \cdot \frac{\triangle \ell_x}{\triangle x} \right)$$

$$= -\frac{1}{\ell_x} \cdot \frac{d\ell_x}{dx}$$

以μ_x表示x歲的死力，這是Gompertz－Makeham 二氏對死力所下之定義。

生命表的計算

①生存數的計算

在計算各年齡的死亡率q_x後，再假設有100,000人口同時出生，便可計算每年有多少人數尚生存著，因為○歲的生存數$\ell_0 = 100,000$。

②死亡數d_x的計算

與生存數的計算法相同，○歲的死亡數$d_0 = 100,000 \times q_0$。

③死亡率q_x的計算

生命表的製作端賴人口普查方能得知，每年的年齡別死亡數都是根據人口動態統計而獲知。死亡率q_x的計算是製作生命表的基礎。計算死亡率，必須有各年齡的人口數值與各年齡的死亡率。

一般是以$m_x = D/P$公式來計算死亡率，m_x叫做中央死亡率（central death rate）是x歲年齡的死亡率。

一年之中的死亡人數與這一年初人口數的比值也就是死亡機率的定義，可以符號表示：

$$q_x = D \div \left(P + \frac{1}{2} D \right)$$

若每一年齡的年中人口數與該年齡到第二年之間的死亡人數已知，便可應用上述公式來計算。但是，這公式所計算的結果，對於二、三歲之前的人口死亡並不準確，尤其是未滿一歲的嬰兒人數，此乃因於嬰兒死亡率低的國家中，前六個月的死亡率往往比後六個月少，所以應當應用下式較為適宜：

$$q_x = D \div \left(P + \frac{4}{5} D \right)$$

二歲以內的死亡率用上式計算誤差會較小。

死亡率m_x及死亡機率q_x的關係

$$\because m_x = \frac{D}{P}$$

$$\therefore P_{mx} = D$$

又

$$q_x = \frac{D}{P + \frac{1}{2}D} \text{ 消去D（以}P_{mx}\text{代入），得}$$

$$q_x = \frac{P_{mx}}{P + \frac{1}{2}m_x} \text{ 消去P，得}$$

$$q_x = \frac{m_x}{1 + \frac{1}{2}m_x}$$

$$= \frac{2m_x}{2 + m_x}$$

注意，在演算時，死亡率m_x的單位是每人而不是每千人。此外，${}_nq_x$ 是表示x 歲到x＋n 歲之死亡機率，以 q_x 表示單一年齡的死亡機率。

④定常人口 T_x 的計算

x 歲以上的總人口以T_x 表示，計算如下：

$$T_x = \int_x^{\infty} \ell_t dt = \sum_{t=x}^{\infty} L_t$$

T_x 是用來表示滿x 歲生存者的人數（ℓ_x）一直到死亡之間的生存年數，又叫做隨後總生存年數（total number of years lived after age X or total after life－time）。

假如考慮到生存數ℓ_x，則計算T_x 時：

$$\because L_x = \frac{1}{2}(\ell_x + \ell_{x+1})$$

故 $$T_x = L_x + L_{x+1} + \cdots\cdots + L_{x+n} + \cdots\cdots$$

$$= \frac{1}{2}\ell_x + \ell_{x+1} + \ell_{x+2} + \cdots\cdots$$

$$=\frac{1}{2}\ell_x+\sum_{n=1}^{\infty}\ell_{x+n}$$

⑤定常人口 L_x 的計算

L_x 是表示 x 歲的人口，計算定常人口 L_x 可用如下公式：

$$L_x=\frac{1}{2}(\ell_x+\ell_{x+1})$$

生存人口延年數，一般稱爲總生存年數（total number of years lived or total life－time between age X and age x＋1 L_x）。L_x 是指滿 x 歲的生存人口 ℓ_x 生存至滿 x＋1 歲之間所生存的總年數。假設滿 x 歲的死亡人平均生存½年，如此：

$$L_x=\ell_{x+1}+\frac{1}{2}d_x$$

$$=\ell_{x+1}+\frac{1}{2}(\ell_x-\ell_{x+1})$$

$$=\frac{1}{2}(\ell_x+\ell_{x+1})$$

依照定常人口 L_x 的定義：

$${}_nL_x=\int_x^{x+n}\ell_t\,dt$$

應用 Euler－Maclaurin 積分公式（求到第二項）計算公式如下：

$${}_nL_x=\frac{n}{2}(\ell_x+\ell_{x+n})+\frac{n^2}{12}(\ell'_x-\ell'_{x+n})$$

$$\ell'_x=\frac{dlt}{dt}\ ,\ t=x$$

依 Lagrange 氏四次挿補公式計算 ℓ'_x：

$$\ell'_x=\frac{1}{12}\ell_{x-2}-\frac{2}{3}\ell_{x-1}+\frac{2}{3}\ell_{x+1}-\frac{1}{12}\ell_{x+2}$$

故　$L_4=\frac{1}{2}(\ell_4+\ell_5)+\frac{1}{12}(\ell'_4-\ell'_5)$

而　$\ell'_4=\frac{1}{12}\ell_3-\frac{2}{3}\ell_4+\frac{2}{3}\ell_6-\frac{1}{12}\ell_8$

$$\ell'_5=\frac{1}{12}\ell_3-\frac{2}{3}\ell_4+\frac{2}{3}\ell_6-\frac{1}{12}\ell_8$$

⑥平均餘命 $\mathring{e}_x$ 的計算

通常計算平均餘命的公式如下：

$$\mathring{e}_x=\frac{1}{\ell_x}\left(1-\int_{y=x}^{\infty} {}_y dl_y\right)$$

一歲以上的平均餘命計算公式：

$$\mathring{e}_x=\frac{1}{\ell_x}\left\{\frac{1}{2}d_x+\left(1+\frac{1}{2}\right)d_{x+1}+\left(2+\frac{1}{2}\right)d_{x+2}+\cdots\right\}$$

$$=\frac{1}{\ell_x}\left\{\frac{1}{2}\Sigma d_x+\Sigma d_{x+1}+\Sigma d_{x+2}+\cdots\cdots\right\}$$

$$=\frac{1}{\ell_x}\left\{\frac{1}{2}\ell_x+\ell_{x+1}+\ell_{x+2}+\cdots\cdots+\ell_w\right\}$$

$$=\frac{1}{2}+\frac{\ell_{x+1}+\ell_{x+2}+\ell_{x+3}+\cdots\cdots+\ell_w}{\ell_x}$$

如此，我們可說 x 歲以上生命表人口累積數也就是 x 歲生存人口尚可活存的延年數，因此，可簡化平均餘命的公式如下：

$$\mathring{e}_x=\frac{T_x}{\ell_x}$$

若使用不同的公式所求出的結果有出入，此乃四捨五入的結果。

⑦死力 μ 的計算

依照定義，死力 μ 的計算如下：

$$\mu_x=\frac{1}{\ell_x}\cdot\frac{dl_x}{d_x}$$

$$=\frac{-\ell'_x}{\ell_x}$$

使用四次補插公式（近似值）求三歲以上者：

$$\ell'_3=\frac{1}{12}\ell_1-\frac{2}{3}\ell_2+\frac{2}{3}\ell_4-\frac{1}{12}\ell_5$$

即 $$\ell'_x = \frac{1}{12}\ell_{x-2} - \frac{2}{3}\ell_{x-1} + \frac{2}{3}\ell_{x+1} - \frac{1}{12}\ell_{x+2}$$

得 $$\mu_x = -\frac{\ell'_x}{\ell_x}$$

以上僅適用於三歲以上死力的計算，二歲以下應用二次補插法計算。

第十章
流行病的統計法則

敏感性及特異性（ Sensitivity And Specificity ）

病患經過診斷，爲肯定的百分比，稱爲敏感性。經診斷，爲否定的百分比，稱爲特異性。敏感性及特異性也就是專門用來表示一種診斷疾病的方法，証明它的效果如何。

至於診斷方法對一疾病是否具有價值，需有如下標準，敏感性＋特異性＞1.0 ，若此二者之和大於2.0 ，則此方法爲佳。如果一診斷方法的敏感性近於1.0 ，說明了這種方法對於未患病的人檢查結果爲陰性，所以，若檢查出呈陽性者便可肯定診斷爲正確的。

下表爲有否疾病的診斷結果：

疾病	診斷		
	陽性	陰性	
有	眞陽性（ TP ）	假陰性（ FN ）	D（＋）
無	假陽性（ FP ）	眞陰性（ TN ）	D（－）
	P（＋）	P（－）	合計（ n ）

敏感性＝ TP／D（＋）　　　疾病盛行率＝D（＋）／n .

特異性＝ TN／D（－）　　　診斷陽性率＝ P（＋）／n .

特異性及疾病流行的程度之試驗方法決定眞陽性反應，就以結核菌素皮膚反應的試驗言，注射後能產生高度敏感反應，若呈陽性反應，則

特異性不高，因其包含過去曾感染者，此時感染但無病徵，部份正在發作的疾病與臨床上活動期之結核病。

可以應用敏感性及特異性來比較二種診斷方法對同一疾病的診斷效果高低，以統計方法評定二種診斷方法的敏感性及特異性的差異意義，可應用 x^2 檢定法。

虛無假說是特異性＋敏感性－1＝0，也就是患病者及無病者的陽性率相同，由所計算出的 x^2 值可求出在臨床診斷時，這虛無假說可以成立，不過，以 Poppler 法診斷，虛無假說不能成立，應捨棄。

有效性（ Effectivity ）

有效性，此一名詞創於 1950 年，是用以表示當時正普遍施行的集團疫苗接種試驗。若以 P_1 代表已接種過疫苗者，以 P_2 表示未接種者，便出現下式：

$$有效性 = 100\,(1 - P_1 / P_2)$$

相對危險性（ Relative Risk ）

研究發生疾病的原因，有二種方法，即集團研究法（ cohort ）與個案對照法（ case－control ），主要的作用是在發現引起疾病的某一成因之危險性的高低。使用集團研究法可以直接估計危險性的高低，個案研究法就不能直接求出答案。

由全體中隨機抽取一部份樣本，便可估算相對危險性的大小，或者是由因素之性質分成有與無二類樣本中求得。個案對照的研究可將疾病分為二類而估得。假設於無病的樣本中隨機抽樣，在個案對照的研究中，產生如下的結果：

	疾病 +	疾病 －	
因素＋	a	c	a＋c
因素－	b	d	b＋d
	a＋b	c＋d	n

$a+b=c+d=\frac{1}{2}n$，在這類研究中，個案組與對照組中之數相等。此 a/b 亦可視爲 P_1/P_2 的合理估計值，而 c/d 爲 P_3/P_4 的合理估計值。因此，

$$\text{相對危險性觀察值}=\hat{\psi}=\frac{ad}{bc}$$

a/b 與 c/d 之比值，可視爲全體相對危險性的估計值，如下：

$$\frac{P_1}{P_2}\div\frac{P_3}{P_4}=\frac{P_1P_4}{P_2P_3}\ (=4)$$

相對危險性估計值的抽樣變異數，可以對數值來表示：

$$\operatorname{var}(\log_e\hat{\psi})=\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}$$

log 爲自然對數(natural or Naperian logarithm)，假使計算時沒有自然對數表，可先以常用對數來計算(以10爲底)，之後再乘上 2.3026 。

假設在一個全體之中，每一分子依某病因分成陽性與陰性，並且依疾病狀態亦分爲陽性及陰性，病因的分類可依對象目前的情形來區分，或依對象過去的病歷來分類，(後者乃是回顧研究法 retrospective study 常用)，疾病的狀態可以某一時期內有否某一現象發生，或某一時刻是否感染了疾病。這三種疾病的分類法一般稱爲盛行 prevalence 與發病 incidence)。

依上述分類法可將全部的研究對象分成如下之二重表(2 × 2 table)(表中的數值是佔全體中的百分比)：

	疾 +	病 −	
因＋	P_1	P_3	P_1+P_3
素－	P_2	P_4	P_2+P_4
	P_1+P_2	P_3+P_4	1

若這些百分比已知，疾病及因素若有相關（asso－ciation）存在，則度量的方法乃依暴露因子的有無來決定染患疾病的危險性之比值。

$$危險性比值 = \frac{P_1}{(P_1 + P_3)} \div \frac{P_2}{(P_2 + P_4)}$$

$$= \frac{P_1(P_2 + P_4)}{P_2(P_1 + P_3)}$$

因為在所有的病因研究中，染患疾病的比率甚小，P_1 比 P_3 小，P_2 比 P_4 小，因此可簡化上式如下：

$$\frac{P_1 P_4}{P_2 P_3} \ (=4)$$

上述的比值也就是近似相對危險性（approximate relative risk），一般都只簡稱為相對危險性（relative risk），此外，還有另一名稱為可能危險性（odds ratio），因為它是 P_1/P_3 及 P_2P_4 的比值，這二個數值可視為代表疾病，英文就是 odds。再者，由於上式中的二個乘積 P_1P_4 及 P_2P_3 是由表中的交叉相乘所得結果，因此也叫做交叉比值（cross－ratio）。

因為相對危險性的估計值是得自許多不同的樣本，因此，要將它們相互組合與比較。有二個比較 2×2 表中的排或欄之不同的百分比，一為 Mantel 與 Haenszel 二者所提出，是個相當簡單的組合方法，這個方法與聯合估計值（pooled estimate）相似。若一 2×2 表中，小分類為 i 時，會產生如下之頻數：

	疾 +	病 −	
因＋	ai	ci	
素－	bi	di	
			ni

則 ψ 的聯合估計值如下：

$$R=\frac{\Sigma(\ a_i d_i / n_i\ )}{\Sigma(\ b_i C_i / n_i\ )}$$

配對（match）的問題是個案對照研究中一個特殊的分類方法，是將每一個案依一重要因素，尋出一對象與之配對，主要目的也就是在比較。每一配對的對象必須成一小樣本，以方便求其相對危險性。之後便可計算 Matel － Haeszel pooled estimate。若有 $\frac{1}{2}$n 的配對數，依其因素之有無，可以形成下列的二重表：

對　照　組

		因素（＋）	因素（－）	
個案組	因素(+)	t	r	a
	因素(−)	s	u	b
		c	d	$\frac{1}{2}n$

上表中的總數是表中各格的頻數，則

$$R=\frac{r}{s}$$

ψ 的對數則為：

$$\log_e(\ P_1 / P_3\ )-\log_e(\ P_2 / P_4\ )$$

＝對數值（有因素時患病的機率）－對數值（無因素時患病之機率）

計算加重平均（weighted mean）之公式如下：

$$\frac{\Sigma wi \ \log \hat{\psi} i}{\Sigma wi}$$

檢定異質性公式如下：

$$x^2_{(9)} = (2.3026)^2 \{ \Sigma wi (\log \hat{\psi} i)^2 - \frac{\Sigma wi \log \hat{\psi} i}{\Sigma wi} \}$$

由以上公式，我們可知個別的估計值中，異質性並不是很顯著，也就是在個別的研究調查中，相對危險性各有不同。

但因爲個別估計值的抽樣變異太大，以致於看不出其確實的差異。若設所有的變異是由於抽樣誤差所引起，則 $\log \hat{\psi} i$ 之變異數如下：

$$\frac{1}{(2.3026)^2 \Sigma wi} = 0.00179$$

第十一章
生命及人口統計均為一種應用統計

一、人口靜態情況的統計

人口密度(The density of population)，是為單位土地面積上所平均居住的人口數量，計算人口密度公式有二：

$$\text{土地面積人口密度}=\frac{\text{年中人口總數}}{\text{土地總面積}}$$

公式〔 11 －1 〕

$$\text{耕地面積人口密度}=\frac{\text{年中人口總數}}{\text{耕地面積}}$$

公式〔 11 －2 〕

基於地形及氣候之因素，公式〔 11 －2 〕所計算出的人口密度，能夠較明顯的顯示土地負擔輕重的情形，但公式〔 11 －1 〕較能將人口密度正確的顯示出。

年齡分配

在生命統計中，年齡的統計是最重要的一環，舉凡出生率、死亡率或編製生命表，或做靜態人口分析都須仰仗年齡分配的資料，此外，一些措施，如經費撥款、衛生設施、教育、兵士預估、社會安全……等計劃亦均端賴人口年齡的分配資料。

年齡分組之組距

①遜巴克（ Sundbrag ）研究人口問題的年齡分組：

年齡 人口增長情形	0～14	15～49	50+
增進型人口	40 %	50 %	10 %
穩定型人口	33 %	50 %	17 %
減退型人口	20 %	50 %	30 %

②威普爾 Whipple 變態人口類型之年齡分組

變態人口分為遷出與遷入二大類，也叫做非正常狀態人口。威普爾認為，在非正常狀態人口中，15～50歲的壯年人口占總人口的比率往往多於或少於50％，與50％之差距愈大，變態也就愈大。壯年人口多於50％以上屬於遷入類型人口，顯示遷入人口多。壯年人口少於50％，便屬於遷出類型人口，顯示遷出的人口多。此乃因遷出遷入的人口大多以15～50歲的壯年人較多，而0～15歲以及50歲以上的人口較少遷出遷入，其影響不大。

③1890年瑞典人口之年齡分組

總共分為0～1歲以下，1～19歲，20～39歲，40～59歲，60歲以上五組。

上述之1890年瑞典之人口及1901年英格蘭與威爾斯之人口，在生命統計上，叫做標準人口，其之所以稱為標準人口，乃係由於上述時期之二國人口，皆為經精確調查的全國人口，並未受移民遷入遷出所影響的常態人口及穩定人口，此種年齡分組法，準確度高，乃為大多數國家普遍應用。

④一般分組法

ⓐ一歲為一組

ⓑ五歲為一組

ⓒ十歲爲一組

一般應用較廣的是以五歲爲一組者，這種方法是高年齡組以下五歲一組，高年齡組以上爲開放性組距。

性別分配

①性比例（ Sex ratio ）

性比例也就是一百女性所當之男性人數，研究生物與醫學發現，孕期中的男胎多於女胎，性比例約爲120左右，但男胎的死亡率較女胎爲高，以致於在出生時的實際比例爲105左右；又出生後的男嬰死亡率又高於女嬰，到了成年之後的性比率就約爲100左右了。又因女性的生命往往比男性長，因此，年齡愈大，性比例也就隨之逐漸降低。

此外，基於宗教及習俗等因素，也可能會影響性比例的高低；經濟狀況較佳、生活環境較好之地區性比例亦可能較低於經濟環境較差、生活環境較劣之地區；戰爭較頻之地區，性比例亦會顯著降低。新開發的地區，男性移民遷入較多，便會使性比例增高。

一般常以下列三種方式來表示性比例：

$$\text{普通性比例}=\frac{\text{該地年中男子總人數}}{\text{該地年中女子總人數}}\times 100$$

(General Sex Ratio)

$$\text{年齡組別性比例}=\frac{\text{該地年中某年齡組男子人數}}{\text{該地年中某年齡組女子人數}}\times 100$$

(Sex ratio by age groups)

$$\text{出生性比例}=\frac{\text{全年出生男嬰數}}{\text{全年出生女嬰數}}\times 100$$

(Sex ratio at birth)

公式〔 11 － 3 〕

上述之普通性比例及年齡組別性比例，皆是以年中靜態的有關資料

爲準，而出生性比例乃以全年的動態資料爲準，所以彼此之間並不完全一樣。

男性、女性占總人口的百分比：

$$\text{男性（女性）占總人口百分比} = \frac{\text{年中男（女）性總數}}{\text{年中總人口數}} \times 100$$

公式〔11－4〕

人口結構（Population Composition）

人口結構乃係將人口依年齡及性別分組後，所顯示出的人口組合情形。若將之繪成圖形，即成人口搭形圖（Age－sex Pyramid），通常有下列四種：

①穩定型：

穩定型的人口是底層的幼年人口及中層之壯年人口相近，老年人口之增減不大。此型態的人口增加較慢，或者根本沒增加，甚至還有減少的可能性，這種型態的形成大多是因爲出生及死亡率都不大的關係。

②金字塔型：

屬於此種形狀的人口是最底層低年齡組的人口較多，而年齡愈大的人口愈少，誠如金字塔型一般，因而名之。常態性質的人口增進大多屬於此型。

③蜂腰型：

顧名思義，此乃如蜂腰一般，中間凹陷，底層與上層較寬廣，也就是幼年及老年的人口都要比壯年人多。這種情形大多發生在戰爭頻繁之國家，或鄰近爲大城市之地區。

④寬腰型：

中年人口最多，多於老年人口與幼年人口，可能屬於此種型態的地區大多爲新興之城市，遷入的人較多之故。

人口平均年齡

一般來說，工業化較顯著之國家，人口平均年齡大多在25歲以上，中度工業化之國家，人口平均年齡約為20歲左右，台灣目前的人口平均年齡約為18歲左右。人口平均年齡的計算，很難求出其算術平均數，這是因為人口年齡分組的高齡組是為開放性不定組距，而以中位數來做為人口平均年齡的代表數值。

由上述的國家工業及人口平均年齡之關係可得此結論：一國人口年齡的高低與其工業化之程度呈反比，高度工業化國家的人口平均年齡往往高於低度工業化的國家。高度工業化國家與落後地區的兒童及老年人社會地位不一，或許有它統計上的因素，前者兒童人口的比例只為後者的一半，因而受重視的程度亦較高。並且前者老年人口的比例竟是後者之四倍，因為在落後地區老年人的比例並不高，因而容易得到晚輩的尊敬與侍候，或者經濟上的支助。高度工業化國家中，每八人中有一老年人，每四人中有一兒童，低度工業化國家（落後地區），15歲以下的人口佔40％，差不多是每2人中卽有一位兒童，每30人中才有一位老年人，約佔3％。

出生率是影響人口平均年齡的最主要因素，若低生育率的情形持續甚久，幼年人口便日益減少，而老年人口日趨增多，人口平均年齡自然逐漸偏高；相反地，若連續一段長時間保持著高出生率，如此，幼年人口便不斷增加，而老年人口逐漸減少，人口平均年齡也就自然要偏低了。死亡率亦會影響人口的平均年齡，但影響要比出生率小多了，因為，若死亡率降低，平均死亡年齡會升高，但幼年的人口亦一年年地成長，況且，我們所指的人口平均年齡並不包括死亡之人口，所以，死亡年齡降低後，若出生率不降低，人口的平均年齡也就要低些了。由此我們得知結論，生育率的下降造成老年人口的增加，生育率的上升導致年青人口之主因。

世界人口年齡結構之發展階段，分為三階段，第一階段為沈重的幼

年依賴者時期(heavy youth dependency)，15歲以下的人口偏高，65歲以上人口偏低，二者的人口即占了總人口的40%以上，這些幼年人及老年人均須依賴15～65歲的壯年人口養育，造成相當之負擔，生產者對依賴人口之比爲3：2。十九世紀末期西歐的人口即爲此情況。第二階段爲輕微的依賴者時期(Light Dependency)，生產者與依賴人口之比，大致爲4：2或5：2。近幾十年來的西歐國家即是此種情況。第三階段爲沈重的高年依賴者時期(Heavy oldage Dependency)，老年人口偏高，幼年人口偏低，幼年人口與老年人口之和和生產年齡的人口差不多，甚或更多者，目前尚無此種情況發生之國家。但自1970年之後，西方的國家已有日漸趨此之勢。

年齡資料正確性的測驗方法：

由於年齡的組合，往往受許多因素的影響，而使得人口年齡的分配成不規則的變化，這不是統計上的錯誤，因此，這些測驗的方法，僅爲參考之用。年齡資料統計正確性的測驗方法大致有下列幾種：

①麥耶爾指數(Myer's index)

麥氏提出十歲以上各年齡組的人口，個位數字相同的各年齡組人口數的總和值是相同的，所以，各個位數相同的年齡人口之加總數字，該是頗近似總人口數之百分之十。實際所得到的百分比與百分之十的差距，其絕對值加總數字，也就是所謂的麥氏指數。

②年齡比例測定法

年齡比例對100的離差的絕對值的平均數也就是年齡比例值，年齡比例值愈小，則表示人口年齡的分配愈適中。年齡比例乃係某一特定年齡組之人口數，對上下相鄰組人口和數平均數所佔的百分比。如果此比例接近100，則表示年齡統計是正確的。

③懷普爾指數(Whipple's index)

$$懷普爾指數=\frac{\Sigma(25,30,35\cdots\cdots 60歲人口數)\times 5}{\Sigma(23至60歲人口數)}\times 100$$

公式〔11－5〕

④聯合國秘書處方法（The united nations Secretarist Method）

聯合國秘書處方法也就是以三倍的性比例值與男女二性個別的年齡比例值的和數來表示，數值愈小，則愈表示人口分配適中。

⑤性比例測定法

性比例是每一百女性人口所當之男性人口數，連續二年齡組性比例相互差數的絕對值的平均數即是性比例測定法中所使用的性比例值。假使人口年齡分配正常，連續二年齡組的性比例變化會很少，不會成不規則變化。通常年齡別性比例的變化，都有固定的型態，一般說來，在青年階段及之前，男性多過女性，性比例因此較高。因爲各年齡組男性的死亡率多高於女性，所以高年齡的女性要比男性多，使得性比例漸次降低。通常60歲以後，性比例低於100，60歲以前，則高於100。性比例值愈低，人口性別年齡也愈適中。

年齡比例測定法、聯合國秘書處方法及性別比例測定數，皆適用五歲分組的資料。

因爲老年人口的年齡組合，常有它較爲特殊的情況，所以，一般都只使用至70歲以下的年齡組合人口，除非有特別的需要，亦可應用至70歲以上。

計算步驟如下：

①求各年齡組性比例

$$某年齡組性比例=\frac{該年齡組男性數}{該年齡組女性數}\times 100$$

②求相鄰二組性比例的差數

可運用上述公式。

③求年齡別比例數

$$\text{男(女)性某年齡組比例數} = \frac{\text{男(女)性某年齡組人數}}{\frac{\text{(較小年齡組人數)}+\text{(較大年齡組人數)}}{2}} \times 100$$

④求各年齡組比例數－100的差數

⑤求性比例值

⑥求年齡比例值

⑦求測驗值

測驗值＝(3×性比例值)＋(男性年齡比例值)＋(女性年齡比例值)

婚姻分配

美國威爾克斯教授(Prof. W. F. Willcox)研究發現，已結婚的夫婦雙方，因有彼此間的鼓勵和照顧，易獲得身心雙方面的調適。並且已婚者的生活有約束，也較規律，支出與收入能斟酌分配使用，生活安定，因此壽命要比未婚者爲高，死亡率亦較低。此外，合法的婚姻是嬰兒出生的根本，由此種種可知，婚姻的分配與生命的現象有極密切的關係。

依人口年齡的分配來說，16歲以上稱爲可婚人口(Marriageable Population)，通常16歲以下的人口約占總人口的⅓左右，也就是說，可婚人口大約爲⅔。

依婚姻的狀況，可婚人口可分爲已婚與未婚二類；已婚人口又可分爲有配偶及無配偶二類；無配偶人口又可分離婚及鰥寡二小類。未婚人口可分爲遲婚(晚婚)、失婚及不適婚三類。失婚的意思是指，適婚，但因某種因素未能找到合適的結婚對象，並且也上了年紀的人。不適婚則是因生理或心理上的缺陷，或者其他不適合結婚等種種因素之人。

可以下表明白顯示人口的婚姻狀況：

可婚人口之婚姻狀況

- 未婚
  - 不適婚
  - 遲婚
  - 失婚
- 已婚
  - 無配偶
    - 鰥寡
    - 離婚
  - 有配偶

二、生命統計及人口統計的依據及重要

人口統計係指運用統計方法，研究人口之種種情況、型態。生命統計乃係運用統計方法，研究人口動態中的種種生命狀況，因此，人口統計及生命統計皆爲一運用統計。

人口統計與生命統計實際上是相類似的，只是在研究上的著眼點略異，人口統計的範圍較生命統計稍爲廣泛些。生命統計的範圍也就是人口統計除去靜態部分，其餘動態部分的出生、死亡、人口增減等。人口統計中的人口密度、人口就業等狀況，看似與生命統計無所關連，但多少總會有些影響的。

生命統計中的標準死亡率的計算及生命表的製作上比起人口統計要詳實些。

茲將人口統計或生命統計之重要性分列於下：

①人口統計資料是政府各種施政方針的依據

一個有爲的政府，必爲人民之教育、建設、經建、社會政策、物資供應、衞生建設……等各種事項進行各項設施或計劃，在擬定各項設施與計劃之前，即須以人口的統計爲其藍本。

②人口政策之擬定有賴人口統計資料之協助

世界人口的增加，端賴人口統計資料的顯示，例如，每年人口增加若干，需有前一年的人口統計資料爲底，知道了世界人口增加的速度，方可策定應變人口急速增加的策略。若人口急速下降，政府亦可研究是否人民生活水準低，衞生保健差，或是是飲食的不均或不足，以便擬定對策。

③以茲工商業界的參考

一個工廠製造財貨，需有多少產量方敷人們使用，不致於發生供不應求或生產過剩的現象，或者投資者在投資經營某項事業之前，也需要用到人口統計表。

④都市計劃之擬訂須以人口統計資料爲本

人口流動的趨勢，可自人口統計資料中，窺出端倪。若在人口統計資料中發現一都市突然湧入大量人口，政府就須研究是否該興建更多的國民住宅供其居住，亦或興建更多的超級市場、購物中心，方便民衆，或者興建交通加強運輸效能。另一方面亦可研判人口外流的因素，加強輔導人口外流地區的各項經建措施，使得一國之內，各地區人口數維持平衡。

取得人口資料的來源有下列五種途徑：

①利用幾何級數來估計

應用此法的首要條件是假設每年人口的增加率固定不變，也就是人口數一年年依幾何級數遞增，步驟是先求出二期之普查人口每年的增加率，其次再求估計年份之人口數。

$$P'_x = P_x(1+r)^n$$

公式〔11－6〕

P'_x：第二次普查年之年中人口數

P_x：第一次普查年之年中人口數

n　：兩次普查期之間隔年數（一般爲十年）

r　：人口增加率

可將公式〔11－6〕化簡爲：

$$\log P'_x = \log P_x + n\log(1+r)$$

$$\log(1+r) = \frac{\log P'_x - \log P_x}{n}$$

$$r = \text{antilog}\frac{\log P'_x - \log P_x}{n} - 1$$

公式〔11－7〕

則：

$$P_t = P_x(1+r)^t$$

公式〔11－8〕

若普查人口不是年中人口，可將公式〔11－6〕及公式〔11－7〕中的n由10改爲分數即可。（普查通常相隔10年，是以n一般爲10）

②按出生、死亡及遷出、遷入之人口數估計

某期估計人口＝（上期普查人口）＋（出生人口）＋（遷入人口）－（死亡人口）－（遷出人口）

公式〔11－9〕

此種估計法較爲繁瑣，因其變動性較大之故，若欲採用此法，須有詳實的人口動態資料登記體制並澈底執行的國家方能有較精確的估計值。

③以圖解曲線來配合估計

將歷年之人口普查資料，以圖的繪製來顯示結果，縱軸為人口數，橫軸為年數。若希望有更精確的估計值，可配合趨勢來進行估計。

ⓐ直線趨勢

$$P'_t = a + bt'$$

公式〔11－10〕

P'_t：估計年期，年中人口數

t'：估計年期

a，b：皆為常數，是自已知的資料中求出：

$$a = \frac{\Sigma P_t}{N}$$

$$= \overline{P}_t$$

$$b = \frac{\Sigma t' P_t}{\Sigma t'^2}$$

公式〔11－11〕

N：年期數

P_t：已知年期人口數

公式〔11－11〕在奇數期時，是以當中間年期t為0，其中間年期以前各期為－1，－2，……之後各期為1，2，……等。在偶數期時，其中間二個年期的t一為－1，另一為1，之前各期為－3，－5，……，之後各期為3，5，……等。

上述公式之目的僅限於明示直線趨勢的配合重點，因人口變動的實際情形不可能成直線趨向，所以，配合圖解趨勢線行人口估計時，可應用到上述公式。

ⓑ以羅吉斯曲線配合來估計人口

羅吉斯曲線(Logistic Curve)公式如下：

$$\frac{1}{P'_t}=k+ab^t$$

公式〔11－12〕

公式〔11－12〕也叫做Pearl－Reed曲線公式。

$$b^{mn}=\frac{\Sigma_3\frac{1}{P_t}-\Sigma_2\frac{1}{P_t}}{\Sigma_2\frac{1}{P_t}-\Sigma_1\frac{1}{P_t}}$$

公式〔11－13〕

$$a=\frac{b^n-1}{(b^{mn}-1)^2}\left(\Sigma_2\frac{1}{P_t}-\Sigma_1\frac{1}{P_t}\right)$$

公式〔11－14〕

$$k=\frac{1}{m}\left[\Sigma_1\frac{1}{P_t}-a\left(\frac{b^{mn}-1}{b^n-1}\right)\right]$$

公式〔11－15〕

m：各部分之項數

n：二次普查之間隔年數

$\Sigma_1\frac{1}{P_t}$，$\Sigma_2\frac{1}{P_t}$與$\Sigma_3\frac{1}{P_t}$，是將$\frac{1}{P_t}$依時間的先後次序分爲相等的三部分，各部分的P_t之倒數和即是$\Sigma_1\frac{1}{P_t}$，$\Sigma_2\frac{1}{P_t}$，$\Sigma_3\frac{1}{P_t}$。

配合羅吉斯曲線來行外插補法估計人口，推計期t項多以3mm－1爲限。

ⓒ二次拋物線

$$P'_t=a+bt'+ct'^2$$

公式〔11－16〕

$$a=\frac{\Sigma t'^4\,\Sigma P_t-\Sigma t'^2\,\Sigma t'^2P_t}{N\Sigma t'^4-(\Sigma t'^2)^2}$$

$$b = \frac{\Sigma t' P_t}{\Sigma t'^2}$$

$$c = \frac{N\Sigma t'^2 P_t - \Sigma t'^2 \Sigma P_t}{N\Sigma t'^4 - (\Sigma t'^2)^2}$$

公式〔11－17〕

此種以外插補法估計人口，僅適用於短期。

④按算術級數來估計

應用此法的首要條件為，假設每年增加的人口數量一致，也就是人口的數量，每年均按算術級數遞增。使用算術級數，須先知其二普查年之年中人口數，再算出二年的人口差數與普查相隔的年數，最後以人口差數除上二普查期之相隔年數，便得知每年人口增加數，如此即可推算年份人口數。

$$P_t = P_x + \frac{t}{n}(P'_x - P_x)$$

公式〔11－18〕

P_t：估計年份之年中人口數

P_x：第一次普查年之年中人口數

P'_x：第二次普查年之年中人口數

n：兩次普查之間隔年數

t：第一次普查期後年數

若普查之期不是在年中，可應用下列化為年中人口：

$$P = h + ad$$

公式〔11－19〕

P：年中人口數

h：普查日人口數

a：七月一日在普查日之後日數（若七月一日在普查日之前，a

爲負數，之後則爲正數）

d：二次普查期間，每天人口增加量

⑤將其他有關因素列入考慮調整估計數字

以上各種估計法所估出之數値若再加上一切足以影響人口增減的各項因素，如交通狀況、人口密度、都市計劃的改變等等，則所計算出的數値便更爲精確了。

三、疾病率及婚姻率的演算

疾病率能影響人口生命力的強衰，與平均年齡的高低，若將疾病率依年齡、性別、職業……各不同的性質來計算，便能澈底了解疾病對於人的威脅。

$$\text{患病頻率} = \frac{\text{全年全人口患病總次數}}{\text{全年人口數}}$$

(Morbidity frequency)

公式〔 11－20 〕

$$\text{患病率} = \frac{\text{染患某種疾病人數}}{\text{年中人口數}} \times 1000$$

(Morbidity rate or attack)

公式〔 11－21 〕

疾病致死率（ Fatality or case fatality ）是一年中染患某種疾病與死於某種疾病的百分比。

$$\text{某種疾病致死率} = \frac{\text{全年死於某種病症人數}}{\text{全年染患某種疾病人數}} \times 100$$

公式〔 11－22 〕

疾病死亡率（ Mortilyty rate ），是年中人口每十萬人染患某種疾病的死亡人數。

$$某種疾病死亡率=\frac{全年染患某種疾病死亡數}{年中人口數}\times 100000$$

公式〔11－23〕

$$對比死亡率=\frac{全年死於某種疾病人數}{全年患病死亡人數}\times 100$$

(Proportionatc mortality)

公式〔11－24〕

$$死產率=\frac{全年死產數}{全年出生人數（包括死產及活嬰）}\times 1000$$

或：

$$死產比=\frac{全年死產數}{全年出生活嬰數}\times 1000$$

公式〔11－25〕

婚姻率（Marriage rate）

出生率足以影響婚姻率的多寡，且婚姻率可顯示出一個社會的狀況。假使，一國之內戰禍不斷、人民失業率激增、經濟蕭條，則結婚率自然降低，相反地，在結婚率高的國家，必是社會安定、人民富庶、豐衣足食。

$$結婚率=\frac{全年婚姻數}{年中人口總數}\times 1000$$

公式〔11－26〕

較合理的計算是以結婚人數爲分子，15 歲以上的可婚人數（包括未婚或已婚喪偶及離婚）爲分母。

$$校正結婚率=\frac{全年結婚人數}{全年可婚人數}\times 1000$$

公式〔11－27〕

離婚率（ Divorce rate ）

$$\text{離婚率}=\frac{\text{全年離婚數}}{\text{年中人口總數}}\times 1000$$

公式〔 11 − 28 〕

離婚必然發生在有配偶的已婚人口中，並且離婚是男女雙方之事，因此，應求校正離婚率（ Correction of divorce rate ）才合宜。

$$\text{校正離婚率}=\frac{\text{全年離婚人數}}{\text{年中有偶人口數}}\times 1000$$

公式〔 11 − 29 〕

人口增進率

①人口指數 Vital index

$$\text{生命指數}=\frac{\text{全年出生人數}}{\text{全年死亡人數}}\times 100$$

公式〔 11 − 30 〕

生命指數係指人口每死亡 100 人後重置的人數，指數大於 100，表示人口增加，小於 100，則表示人口減少。近年來，無論何處，人口指數均無小於 100 者，因此生命指數在生命統計上很少使用。

②人口的自然增長率 Nature increase rate 及推行計劃生育的途徑

$$\text{自然增長率}=\frac{\text{全年出生數}-\text{全年死亡數}}{\text{年中人口數}}\times 1000$$

或者：

$$\text{自然增長率}=\frac{\text{全年出生數}}{\text{年中人口數}}\%-\frac{\text{全年死亡數}}{\text{年中人口數}}\%$$

$$=\text{粗出生率}-\text{粗死亡率}$$

公式〔 11 − 31 〕

經調查，台灣地區的人口增長率，雖逐年有降低的趨勢，但若考慮本區的土地面積、人口密度與耕地面積，比起其他國家地區，人口密度可謂世界之冠，因此，再降低人口增長率是絕對必要的。首先，節制生育是十分重要的一環，推行家庭計劃，減低出生率，注重質的培養，勿使量再增加。

至於人口未來增長的趨向，可以自然增加率來求出人口倍增所需之年數，也就是自然增加的百分率除七十，誤差不到百分之一。

人口增加，即形成年幼者多於壯年者與老年者，年幼者屬於依賴人口，壯年者須耗費許多精力、時間與金錢於年幼之嬰幼兒身上，自嬰兒時期的哺乳眷顧，學齡時期的養育教育，畢業之後的就業或婚姻問題都是一家庭的負擔，因之，世界上較先進之工業國均極力倡導家庭計劃，使新生人口降低，日本在二次大戰結束時出生率爲千分之卅五，如今已降至半數，大約爲千分之十七左右，人口自然增加率僅爲千分之五。反觀台灣地區，若粗出生率能降至千分之十五，自然增加率降至千分之十，人民的平均所得勢必會高出許多。

印度前總理尼赫魯曾公開表示：「倘使我們國家的人口，只爲現在人口之半，則我們印度將會是個更進步的國家。」今天，亞非部份貧困地區，仔細研判，可發現貧窮、落後之國家，常是高出生率的國家。根據專家的估測，全世界中約有⅔的人口，缺乏足夠之糧食以維持足夠的體力與應有的健康。若依中等速率計算人口之增長，到了西元二〇〇〇年，人口估計將超過七十億，約莫等於現在的二倍之多，可想像屆時人們的生活是如何的困苦了。

降低出生率，可由下列幾項要點着手：

①提高人們的智識，認清高出生率對我們本身、個人或國家的危害有多大，多產家庭對子女、父母、整個家庭有多少不良之影響，之於國家社稷又有多大之妨礙。爲了個人的精神、物質生活，爲了國家的負擔，

爲了整個世界人口壓力的緩和，降低人口出生率是刻不容緩之務。此可經由大衆傳播媒體，如電視、收音機廣播、教育之功能以協助之。

②善用有效時機，協助避姙，例如各城市鄉村在舉行各項聚會時，加強宣導避姙之方法，認識節育的重要性。

③採取半強硬政策，明定子女數，以每個家庭二位子女爲限，或獎勵已有二個孩子的夫婦進行結紮手術，手術費用全勉，並發予獎勵金。

④撥發足夠的經費，以辦好節制生育的各項宣導活動或家庭計劃措施。衞生行政機構專設一計劃生育的單位，負責推行節育的工作，此項經費是省不得的。美國前總統詹森曾說過：「在節制人口方面投資不到五元，可以產生相當於100元投資在發展經濟的價值。」

⑤嚴格執行婚姻法，限制未滿廿歲的青年男女不得結婚，因未滿廿歲的青年或青少年，身心發展未臻成熟，結婚後容易發生各種問題，早婚生下的子女也可能是不健康的兒童。查先進國家的結婚年齡，女子大多在25歲以上，男孩多在30歲左右。因此，爲了個人的身心健康、事業前途，爲了家庭的穩定與幸福，爲了子女的心智成長，禁止早婚的發生是合理的。

⑥在各地的衞生機構內，設專任護士，輔導新婚夫婦，使用避孕方法及免費供應避孕藥物，使其需要避孕的夫婦均能方便及充足的獲得並確實施行。

四、出生率的意義及計算

出生率(Birth rate)

全年所出生的活嬰兒，所占年中全人口的千分比，即爲出生率，也就是年中每千人口全年出生的活嬰兒數。(這裏不包括死產嬰兒)

出生率在15‰到25‰之間，稱爲低出生率，25‰到35‰之間稱

爲中度出生率，35 %到45 %之間稱爲高出生率。大致上，各地區的人口出生率都介於15 %至50 %之間，很少有低於15 %或高於50 %者。人口的出生率，除了和人口的變動、增減有關外，且與國民水準及經濟、教育均有關。一般說來，鄉村地區的出生率高於城市地區，落後地區的出生率高於先進工業國，所得及文化水準較低的國家出生率高於所得及文化水準較高的國家，這是必然之結果。

魏曾(W. F. Wertheim)以百分之四十審查法來解釋出生率與人口年齡關係，十四歲以下之人口占總人口百分之四十，則出生率至少在千分之四十以上。不論變化如何的小，只要是在千分之四十上下變動，十四歲以下占人口總數的百分數，定也隨着變動。

粗出生率(Crude birth rate)，亦有稱之爲普通出生率者(General birth rate)。

$$粗出生率=\frac{全年出生活嬰數}{年中人口總數}\times 1000$$

公式〔11 − 32〕

生育率

出生率只能表示人口出生的大概，不能表示人口生育的能力。以女子人數爲基準，計算出生率，則爲生育率，此乃因生育僅以女子爲主體之故。

$$一般生育率=\frac{全年出生活嬰數}{年中婦女總人數}\times 1000$$

公式〔11 − 33〕

$$普通生育率=\frac{全年出生活嬰數}{年中15至45歲育齡婦女人數}\times 1000$$

公式〔11 − 34〕

$$年齡組別生育率=\frac{該年齡組婦女全年出生活嬰數}{年中該年齡組婦女總數}\times 1000$$

公式〔 11 － 35 〕

一般說來，年輕婦女生育力強，年齡較大者，生育力較弱，因此，若需更臻詳細分析婦女出育狀況，可應用年齡組別生育率（ Age – Specific fertility rate ）。

總生育率＝15 至45 歲各年齡組別婦女生育率的總和

公式〔 11 － 36 〕

由公式〔 11 － 36 〕可得知，總生育率（ Total fertility rate ）也就是婦女年齡別生育率的總和。

婚姻生育率（ Married fertility rate ）也可叫做配偶生育率（ Lagitimat fertility rate ），是以已婚、有配偶的育齡婦女做爲計算生育率的基準，所求出的只是合法子女的生育率，因爲那不包括非婚生的子女。

$$婚姻生育率=\frac{全年已婚有偶育齡婦女生育活嬰數}{年中某年齡組已婚有偶婦女數}\times 1000$$

公式〔 11 － 37 〕

某年齡組別婚姻生育率

$$=\frac{某年齡組全年已婚有偶婦女生育活嬰數}{年中某年齡組已婚有偶婦女數}\times 1000$$

公式〔 11 － 38 〕

繁殖率（ Reproduction rate ）

上述的各項生育率，僅就一代婦女本身而言，若要計算出下代人口的繁殖（ Reproductivity ）或再生（ Reproduction ），便需以繁殖率（ Reproduction rate ）來計算。

淨繁殖率（ Net Reproduction rate ）

除去生命表中育齡期間婦女死亡人數，以實際活著的育齡婦女人數來計算繁殖率，稱爲淨繁殖率。

粗繁殖率（ Gross Reproduction rate ）

$$粗繁殖率 = \Sigma \frac{全年某年齡組生育活女嬰數}{年中某年齡組婦女人數} \times 1000$$

公式〔 11 − 39 〕

粗繁殖率是指1000育齡婦女在15到45歲30年中育齡期間所生育活女嬰數，也就是將總生育率中的男嬰數扣除後，所計算出的生育率。

計算粗繁殖率，是假定每一女嬰都能活到育齡，也就是 45 歲以上，但此假定是不可能的，因此，計算上常有偏差。

死亡率（ death rate ）

①粗死亡率（ Crude death rate ），亦叫做普通死亡率（ General death rate ），公式如下：

$$粗死亡率 = \frac{全年死亡人數}{年中人口總數} \times 1000$$

公式〔 11 − 40 〕

死亡率的高低，與出生率的高低有一定的關係。一般情況下，粗死亡率很少超過35％，或低於8％。若出生率介於25％與35％之間，亡率大多會高於10％或15％。

死亡率的計算分爲屬人與屬地兩種。屬人主義，是指設籍某地常住之人口，不論其死於何地，皆屬於常住地區的死亡人數。屬地主義，是指死亡者若於A行政區內死亡，不論其設籍何處，皆屬於死亡所在地之死亡人數。

死亡率若逐年降低，則表示該國的醫藥及保健設施完善，經濟、社會進步，政治、人民生活安定。

②標準化死亡率（ Standardized death rate ），是以標準人口

的性別年齡分配爲依據，將不同的年齡、性別分配地區使之爲同一基礎，除去人口組成的差異，純比較死亡率。

因爲各地區的人口結構，年齡、性別分配各有不同，因此不能直接由粗死亡率的多寡來評定該地區的生活環境、衞生及文化水準。以性別來說，男子的死亡率往往高過女子，以年齡來說，嬰兒及老年人的死亡率都要比年青人高，若僅考閱粗死亡率，會影響其正確程度。

爲使死亡率，不受年齡及性別組合的影響，須以標準化死亡率來計算。

ⓐ標準化死亡率的間接計算

若僅知一地死亡總數及男女各年齡組人數，須應用間接法計算標準化死亡率。因爲沒有當地男女性年齡別死亡率，只好以標準人口的性別年齡的特殊死亡率來代替，如此，要計算標準化死亡率要有如下三步驟，因此稱之爲間接法。

1 先求指標死亡率(Index death rate)

$$\text{指標死亡率}=\frac{\Sigma\left(\begin{matrix}\text{標準人口性別}\\\text{年齡別死亡率}\end{matrix}\times\begin{matrix}\text{當地人口性別}\\\text{年齡別分配}\end{matrix}\right)}{\Sigma\left(\text{當地人口性別年齡別分配}\right)}\times 1000$$

或

指標死亡率＝

$$\frac{\text{按標準人口性別年齡別死亡率計算的當地人口死亡數}}{\text{當地人口總數}}\times 1000$$

公式〔 11 － 41 〕

2 再求標準化因子 (Standardizing factor)

$$\text{標準化因子}=\frac{\text{標準人口的普通死亡率}}{\text{指標死亡率}}$$

公式〔 11 － 42 〕

標準人口之普通死亡率及指數死亡率差異的大小，要看當地人口結

構的年齡與性別組合和標準化人口組合相似的程度，若二者很近似，標準化因子則為一，否則不能為一。

3. 最後求標準化死亡率

某地標準化死亡率＝當地人口普通死亡率×標準化因子

或

$$某地標準化死亡率＝當地人口普通死亡率\times\frac{標準人口之普通死亡率}{指標死亡率}$$

公式〔11－43〕

假如標準化因子中的分子「標準人口普通死亡率」與分母「指標死亡率」相同，二者之比率即為1，表示標準人口的年齡、性別組合與當地人口組合一樣。若分母的當地「指標死亡率」小於分子的「標準人口普通死亡率」，則標準化因子大於1，表示當地人口，青少年所佔比例大於標準人口，有使當地普通死亡率偏低的趨勢，所以須要提高。

若分母的指標死亡率大於分子的標準人口普通死亡率，標準化因子便小於1，表示當地人口組合中的老年人所佔比例大於標準人口，有使當地普通人口死亡率偏高的趨勢，所以要調整降低。

調整因人口性別年齡組成不同所附於死亡率的影響，即是標準化因子的功能。標準化因子的計算，須依靠戶口普查所得到的當地年齡性別分組的資料。

ⓑ標準化死亡率的直接計算

$$標準化死亡率=\frac{\Sigma\left(\begin{matrix}標準人口的性別\\年齡組別人數\end{matrix}\times\begin{matrix}當地相同性別年齡\\組別的特殊死亡率\end{matrix}\right)}{\Sigma(標準人口之性別年齡組別人數)}\times 1000$$

或

$$標準化死亡率 = \frac{依當地年齡性別組別計算的標準人口死亡總數}{標準人口總數} \times 1000$$

公式〔11－44〕

計算不同地區死亡率的比較的標準化死亡率，不一定要選定特別的標準人口，也可以僅以其中一個地區的人口性別年齡分配作為標準。計算標準化死亡率，可將標準人口化為百萬計算。

③特特死亡率（ Specific death rate ），若要研究特殊情況的死亡率，例如不同年齡組別死亡率、職業類別死亡率、婚姻狀況死亡率、各種疾病死亡率……等等，不能應用粗死亡率，而須以特殊死亡率來計算。

$$某類特殊死亡率 = \frac{全年某種特殊情況死亡數}{年中某種特殊情況總人數} \times 1000$$

公式〔11－45〕

ⓐ性別、年齡的死亡率

$$男性（或女性）某年齡死亡率 = \frac{全年某年齡之男性（或女性）死亡數}{年中某年齡之男性（或女性）人數} \times 1000$$

公式〔11－46〕

ⓑ嬰兒死亡率（ lnfant mortality rate ）

$$嬰兒死亡率 = \frac{全年一歲以下嬰兒死亡數}{全年出生活嬰總數} \times 1000$$

公式〔11－47〕

按英格蘭威爾士死亡統計的經驗，一年中死亡嬰兒數之百分之七十，可能為當年出生，另百分之三十，可能是上一年出生者，因為一歲以

內死亡的嬰兒，並不一定皆爲當年出生者。因此公式〔11－46〕可校正爲：

$$\text{校正嬰兒死亡率} = \frac{\text{全年一歲以下嬰兒死亡數}}{\frac{7}{10}(\text{當年出生活嬰數}) + \frac{3}{10}(\text{上年出生活嬰數})} \times 1000$$

公式〔11－48〕

通常嬰兒死亡率和總人口死亡率之間，有一種相互的關係，高者，二者皆高，低者，二者均低。出生後幾天死亡的嬰兒常未辦理出生與死亡登記，所以嬰兒死亡率都偏低了。一般，一個月的死亡數比之後的五個月多，而前五個月的又比後六個月的死亡數多，出生未滿一個月（一般以 28 天計）死亡者，稱爲幼嬰死亡率（Neo－natal mortality rate）。

$$\text{幼嬰死亡率} = \frac{\text{全年出生不滿一月死亡數}}{\text{全年出生活嬰數}} \times 1000$$

公式〔11－49〕

或

$$\text{幼嬰死亡比} = \frac{\text{某月出生不滿一月之嬰兒死亡數}}{\text{某月出生活嬰數}} \times 1000$$

公式〔11－50〕

ⓒ其他特殊死亡率

其他特殊死亡率，如職業別死亡率、都市別生亡率、婚姻別死亡率……都是特殊死亡率。一般說來，工人的死亡率高於職員，鄉村的死亡率低於都市，未婚的死亡率高於已婚。

五、從業情況的種類及分析

不管是在生命統計上，或是人口統計上，一個國家，其人民就業的狀況都是相當重要的。以生命統計來說，由統計的資料中可知，工業社會的出生率大都比農業社會的出生率低，因此，人口增加率大多與工業化的程度呈反比的狀態，農村地區人民的壽命一般都高於工業地區，這些資料都顯示出來就業的情形與國民的健康、壽命有絕對的關係。以人口統計來說，由人民就業情形可分析出人力運用的情形及社會經濟的狀況，可作爲擬定社會及經濟發展計劃的方針。

①行業(Industry)與職業(Occupation)的分類

行業，係指工作者所屬的經濟活動部門。職業，係者工作者個人本身所擔任的職務或工作。此二者的意義是不同的，所以，每一行業有它主要的經濟活動，由於分工的因素，需要各種不同性質的工作人員，同一職業的工作人員，常在各不同的行業中。

一般分行業爲下列三類：

第一類——農林漁牧業。

第二類——礦業、製造業、電器業等。

第三類——商業、運輸業、服務業等。

依我國之「台灣職業分類典」，將職業分爲十大類，於每一大類中又分若干中類，於每一中類中又分若干小類，又於每一小類中分若干細類。

此十大類爲：

第○類——專業性、技術性職業人員及其有關人員。

第一類——行政及管理人員。

第二類——佐理人員。

第三類——買賣工作人員。

第四類——農夫、漁夫、獵人、伐木人員及其有關人員。

第五類——礦工、採石工及其有關人員。

第六類——運輸及交通工作人員。

第七、八類——技術工匠、生產程序工及不屬於他類之體力工作人員。

第九類——服務、運動及娛樂之工作人員。

雖然農業的技術一直在不斷的進步，農作物的收穫量也愈來愈豐富，但由於全世界都普遍地走入工業化的社會，從事農業生產的人日益減少，自第一類的農業行業趨向第二類及第三類行業，並且第二類行業的增加比第三類大。以美國爲例，1850 年時，美國農人占全部經濟活動人口的65%，1870年時，降至53%，1900年時，又降至 38%，到了1950年，僅剩13%了。

以職業來說，在先進的工業化國家中，專業性、技術性及其有關人員占所有勞動力的5%～8%之間，較未臻工業化之國家，約在2%以下。

② 從業狀況的分類

依據人口是否屬於經濟性活動從業人員，可分爲勞動力人口與非勞動力人口二類，其中，勞動力人口又可分爲平民勞動力與武裝勞動力二類，平民勞動力是指參與生產工作者，武裝勞動力是指保國衞民的軍人。勞動力，一般是指在標準調查週中，年滿十五歲，具有工作能力與工作志願的普通百姓。通常我們所謂的勞動力都是指平民勞動力。

依人口是否屬於經濟性從業標準，將台灣省社會處勞動力調查所分人口類別分述如下：

ⓐ勞動力

1 就業人口：於標準調查週內，從事有酬勞的工作一個小時以上，或者每週十五小時以上的無酬家屬工作。

(1)完全就業人口：就業人口中除去不完全就業人口卽爲完全就業人口。

(2)不完全就業人口：於調查標準週內，因健康情形欠佳、或功課太忙、家務繁多、天氣影響……等因素，工作未滿 36 小時者。

依從業身分，就業人口可分下述四類：

(1)業主（Employers）：爲自己工作，並雇有助手人員。

(2)受僱人員（Employees）：受雇於人，支領薪水者，又可分私人雇用與在政府及公立學校與公營事業單位任職二類。

(3)自營作業者（Workers on own account）：爲自己工作，無雇用助手者。

(4)無酬工作之家屬（Vnpaid family workers）：在家人經營的營利企業中任職，但並不支領薪水。

2. 失業人口：在標準調查週內，有工作能力與工作意願，未能獲得有酬工作一小時以上或每週十五小時以上無酬家屬工作，而有尋找職業的行爲者。

(1)初次尋職者。

(2)非初次尋職者：因所任職之工廠、公司歇業，或技能不當，健康欠佳，或工作志趣不合，自動請辭者。

ⓑ勞動力以外之人口

1. 潛在勞動力：十五歲以上，具有工作能力，但却無工作意願或暫時無法工作的無業者，諸如，在學學生，準備繼續升學者，暫時不打算工作者，家務太忙者。

2. 預備勞動力：十五歲以下的兒童，可稱爲預備勞動者。

3. 無法勞動人口：身心缺陷、老年人，永久不能工作的人口。

③ 勞動力人口需求的估計法

勞動力人口需求的計算，不容易有絕對的標準，此乃因涉及人力需求的各種因素太多了，因此只能大概地作個估計，雖不能達到百分之百的絕對精準，但亦有相當之價值。

ⓐ某行業人力需求的估計法

1. 預計生產量——生產之始，應先有生產量的預計，依此生產量之目標從事生產工作。計劃生產量可依國民生產毛額與國民的消費水準，或是輸出輸入量之多寡，資金的是否充足，原料的供應是否足夠等情形來決定。

2. 預計生產力——一般說來，生產力是祇有提高，少有降低的，此乃因生產技術的進步及人力素質提高的關係。

3. 預計工作時數——未來勞力的需求與工作時數的多寡，有絕對的關係。由於不完全就業的人口逐漸減少，事實上，有效的工作時數，也都是有增無減的。

決定了以上三項估定數值後，便可利用公式以求出所需的勞動力人數。

$$生產力=每人時之生產量=\frac{總產量}{總人時數}$$

$$總人時數=\frac{總產量}{每人時之產量}$$

$$總人數=\frac{總人時數}{每人工作時數}$$

公式〔 11 － 51 〕

ⓑ某職業需求人數估計法

1. 以已有之資料，計算每類職業人數，同時計算其百分比。若無同類資料，可以較進步的工廠代替，以所調查出的數值作爲同

類職業的標準。

2. 考慮各種可能變化的因素，以校正自前一步驟求出的比例。

3. 將各類職業就業的人數乘以修正率，便是各類職業的需求人數。

4. 將各行業中，同一職業的數字相加，便是該類職業的總需求人數。

④一國人口的平均工作年數的計算

此項計算可以人口年齡資料及生命表爲根據。依照人的一生能工作多少年數，分爲二種，一爲毛工作年數（Gross years of active life），另一爲淨工作年數（Net years of active life）。毛工作年數是取各不同年齡組的經濟活動人口數，分別占各該組總人口數的比例，可簡稱爲「經濟活動人口比率」，以此乘上各組的組距，則爲各組的總人口平均每人在該一年齡時期內的毛工作年數，再以此每人毛工作年數相加，便是平均每人一生的毛工作年數。

淨工作年數是應用生命表中的「生存人口數」（ℓ_x）與「靜止人口數」（${}_nL_x$）二種資料來計算，這種計算較爲麻煩，要先用各年齡組的靜止人口數，乘上該年齡組的經濟活動人口比率，求出該一年齡組的靜止人口在該組內的工作總年數，其次，再將各個年齡組中的靜止人口工作總年數，從最後的一組向前累計，在各年齡組內的累計數就是靜止人口在各該年齡之後的工作年數累計數，所以年齡愈低的一組，累計數便愈大，第三步驟，以此累計數除以生存人口數，便爲各年齡組中的人口，平均每人在以後的淨工作年數，年齡愈高的一組，其每人平均在某一年歲以後的工作年數，也就愈少。

若要計算更精確的平均每人工作年數，要以上述之平均每人工作年數所用之經濟活動人口比率，以有業人口比率代之。一般所稱的經濟活動人口均包含失業人口在內，但因失業人口具有鞭策有業人口更加努力工作的能力，以促進生產事業，所以也列入經濟活動人口之一部分，因

此，在計算平均每人一生之工作年數時，以經濟活動人口爲依據。

人口之教育程度

分析人口之教育程度與生命統計有很大的關係，並且可顯示出人力資源的素質情形。觀察嬰兒的出生率及死亡率，可知其與父母知識的程度成反比，也就是說，知識程度較高的父母，其子女的死亡率較低，反之則較高。

計算人口教育程度，可利用以下公式：

$$\text{及齡兒童就學率}=\frac{\text{六歲兒童就學數}}{\text{六歲兒童總數}}\times 100$$

公式〔11－52〕

$$\text{學齡兒童就學率}=\frac{6\sim12\text{ 歲兒童就學數}}{6\sim12\text{ 歲兒童總數}}\times 100$$

公式〔11－53〕

由於我國之國民教育自五十七學年度延長至九年，因此可改上二式爲：

$$\text{及齡兒童就學率}=\frac{6\text{ 歲、}13\text{ 歲人口就學數}}{6\text{ 歲、}13\text{ 歲人口總數}}\times 100$$

公式〔11－54〕

$$\text{學齡人口就學率}=\frac{6\sim15\text{ 歲人口就學率}}{6\sim15\text{ 歲人口總數}}\times 100$$

公式〔11－55〕

$$\text{某級學校畢業生升學率}=\frac{\text{升學入學人數}}{\text{某級學校應屆畢業生人數}}\times 100$$

公式〔11－56〕

$$\text{文盲率}=\frac{12\text{ 歲或 }15\text{ 歲以上文盲數}}{12\text{ 歲或 }15\text{ 歲以上人口數}}\times 100$$

公式〔11－57〕

$$\text{現住人口之某級學校在學學生比率} = \frac{\text{某級學校學生數}}{\text{全年齡人口總數}} \times 100$$

公式〔11－58〕

現住人口之某年齡以上的教育程度比率

$$= \frac{\begin{matrix}\text{某年齡以上之某級學校畢業} \\ \text{及肄業人數（未再升學者）}\end{matrix} + \begin{matrix}\text{某級學校某年齡} \\ \text{以上在學學生數}\end{matrix}}{\text{某年齡以上人口數}} \times 100$$

公式〔11－59〕

第十二章
抽樣調查的方法及應用

一、平均之抽樣誤差

平均的抽樣誤差(The Sampling Error of A Mean)

若在研究調查中，全體的各個數值爲未知數，然其平均值與散布程度也就無法得知。只有不斷地在全體中抽取樣本，計算它平均值的平均 $\mu\bar{x}$ (mean of sample means)來推定 μ，方能求得樣本平均值 $\bar{x}$ 與全體平均值(μ)一致。

減小抽樣變異(sampling variation)需考慮下述二點：

①所抽樣之全體本身的變異。

②抽樣的大小。

若全體的變異大，樣本數小，如此，抽樣平均的差異勢必較大。由理論上說明如何自抽樣的樣本特性推斷全體的特性。假設某一全體由 a、b、c、d 四個數所構成，現自其中隨機抽樣，每次只取出一個數，登記之後，再放回原處，然後另再抽出一數，同樣登記下後，再放回原處，依此不斷的重複，抽出、放回、抽出、放回，以二個數爲一組，樣本大小卽爲2，如此，其組合的情形便可能有如下 16 種：aa，ab，ac，ad，ba，bb，bc，bd，ca，cb，cc，cd，da，db，dc，dd。將這 16 組的平均值視爲一組頻數分布，可求出其平均值及標準偏差。

S^2	組別	$\overline{X}$	$\overline{X}-\mu_{\overline{x}}$	$(\overline{X}-\mu_{\overline{x}})^2$
0	a,a	$\frac{1}{2}(a+a)$	$\frac{1}{2}(a+a)-\mu$	$\frac{1}{4}(2a-2\mu)^2$
$\frac{1}{2}(a-b)^2$	a,b	$\frac{1}{2}(a+b)$	$\frac{1}{2}(a+b)-\mu$	$\frac{1}{4}[(a+b)-2\mu]^2$
$\frac{1}{2}(a-c)^2$	a,c	$\frac{1}{2}(a+c)$	$\frac{1}{2}(a+c)-\mu$	$\frac{1}{4}[(a+c)-2\mu]^2$
$\frac{1}{2}(a-d)^2$	a,d	$\frac{1}{2}(a+d)$	$\frac{1}{2}(a+d)-\mu$	$\frac{1}{4}[(a+d)-2\mu]^2$
$\frac{1}{2}(a-b)^2$	b,a	$\frac{1}{2}(b+a)$	$\frac{1}{2}(b+a)-\mu$	$\frac{1}{4}[(b+a)-2\mu]^2$
0	b,b	$\frac{1}{2}(b+b)$	$\frac{1}{2}(b+b)-\mu$	$\frac{1}{4}(2b-2\mu)^2$
$\frac{1}{2}(b-c)^2$	b,c	$\frac{1}{2}(b+c)$	$\frac{1}{2}(b+c)-\mu$	$\frac{1}{4}[(b+c)-2\mu]^2$
$\frac{1}{2}(b-d)^2$	b,d	$\frac{1}{2}(b+d)$	$\frac{1}{2}(b+d)-\mu$	$\frac{1}{4}[(b+d)-2\mu]^2$
$\frac{1}{2}(a-c)^2$	c,a	$\frac{1}{2}(c+a)$	$\frac{1}{2}(c+a)-\mu$	$\frac{1}{4}[(c+a)-2\mu]^2$
$\frac{1}{2}(b-c)^2$	c,b	$\frac{1}{2}(c+b)$	$\frac{1}{2}(c+b)-\mu$	$\frac{1}{4}[(c+b)-2\mu]^2$
0	c,c	$\frac{1}{2}(c+c)$	$\frac{1}{2}(c+c)-\mu$	$\frac{1}{4}(2c-2\mu)^2$
$\frac{1}{2}(c-d)^2$	c,d	$\frac{1}{2}(c+d)$	$\frac{1}{2}(c+d)-\mu$	$\frac{1}{4}[(c+d)-2\mu]^2$
$\frac{1}{2}(a-d)^2$	d,a	$\frac{1}{2}(d+a)$	$\frac{1}{2}(d+a)-\mu$	$\frac{1}{4}[(d+a)-2\mu]^2$
$\frac{1}{2}(b-d)^2$	d,b	$\frac{1}{2}(d+b)$	$\frac{1}{2}(d+b)-\mu$	$\frac{1}{4}[(d+b)-2\mu]^2$
$\frac{1}{2}(c-d)^2$	d,c	$\frac{1}{2}(d+c)$	$\frac{1}{2}(d+c)-\mu$	$\frac{1}{4}[(d+c)-2\mu]^2$
0	d,d	$\frac{1}{2}(d+d)$	$\frac{1}{2}(d+d)-\mu$	$\frac{1}{4}(2d-2\mu)^2$

a、b、c、d全體的平均值μ與標準偏差S的計算如下：

$$\mu=\frac{1}{4}(a+b+c+d)$$

$$\sigma=\sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

$$=\sqrt{\frac{(a-\mu)^2+(b-\mu)^2+(c-\mu)^2}{4}}\ +(d-\mu)^2$$

$$=\sqrt{\frac{a^2+b^2+c^2+d^2-2\mu(a+b+c+d)+4\mu^2}{4}}$$

$$=\sqrt{\frac{a^2+b^2+c^2+d^2-4\mu^2}{4}}$$

由16組的平均值$\bar{x}$所重新組合成的平均值爲$\mu\bar{x}$（mean of sample means），標準偏差$\sigma\bar{x}$與樣本變異數的平均值μ_s^2，計算於下：

$$\mu\bar{x}=\frac{1}{16}\Big(\frac{a+a}{2}+\frac{a+b}{2}+\frac{a+c}{2}+\frac{a+d}{2}$$

$$+\frac{b+a}{2}+\frac{b+b}{2}+\frac{b+c}{2}+\frac{b+d}{2}$$

$$+\frac{c+a}{2}+\frac{c+b}{2}+\frac{c+c}{2}+\frac{c+d}{2}$$

$$+\frac{d+a}{2}+\frac{d+b}{2}+\frac{d+c}{2}+\frac{d+d}{2}\Big)$$

$$=\frac{1}{16}\left[\frac{8(a+b+c+d)}{2}\right]$$

$$=\frac{1}{4}(a+b+c+d)$$

$$=\mu$$

$$\sigma_{\bar{x}}^2 = \frac{\Sigma(\bar{x} - \mu_{\bar{x}})^2}{N}$$

$$= \frac{1}{16} \times \frac{1}{4} [(2a - 2\mu)^2 + (2b - 2\mu)^2 + (2c - 2\mu)^2 + (2d - 2\mu)^2 + 2(a + b - 2\mu)^2 + 2(a + c - 2\mu)^2 + 2(a + d - 2\mu)^2 + 2(b + c - 2\mu)^2 + 2(b + d - 2\mu)^2 + 2(c + d - 2\mu)^2]$$

$$= \frac{1}{16} \{ [(a^2 + b^2 + c^2 + d^2) - 2\mu(a + b + c + d) + 4\mu^2] + \frac{1}{2} [3 \times (a^2 + b^2 + c^2 + d^2) + 2(ab + ac + ad + bc + bd + cd) - 4\mu \times 3(a + b + c + d) + 6 \times 4\mu^2] \}$$

$$= \frac{1}{16} \{ [(a^2 + b^2 + c^2 + d^2) - 4\mu^2] + \frac{1}{2} [2(a^2 + b^2 + c^2 + d^2) + (a + b + c + d)^2 - 24\mu^2] \}$$

$$= \frac{1}{16} [2 \times (a^2 + b^2 + c^2 + d^2) - 8\mu^2]$$

$$= \frac{1}{2} \times [\frac{(a^2 + b^2 + c^2 + d^2) - 4\mu^2}{4}]$$

$$= \frac{1}{2} \times \sigma^2$$

$\because n = 2 \qquad \therefore \sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$

即 $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$$\text{樣本變異數 } S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$\text{其平均爲 } \mu S^2 = \frac{\Sigma(S^2)}{N}$$

$$\begin{aligned}
&= \frac{1}{16} \times [(a-b)^2+(a-c)^2 \\
&\quad +(a-d)^2+(b-c)^2 \\
&\quad +(b-d)^2+(c-d)^2] \\
&= \frac{1}{16}[3(a^2+b^2+c^2+d^2) \\
&\quad -2(ab+ac+ad+bc \\
&\quad +bd+cd)] \\
&= \frac{1}{16}[4(a^2+b^2+c^2+d^2) \\
&\quad -(a+b+c+d)^2] \\
&= \frac{1}{16}[4(a^2+b^2+c^2+d^2) \\
&\quad -16\mu^2] \\
&= \frac{1}{4}[(a^2+b^2+c^2+d^2)-4\mu^2] \\
&= \sigma^2
\end{aligned}$$

由上述之計算，可得以下之三項結論：

① $\mu\bar{x} = \mu$，即樣本平均值 $\bar{x}$ 的分布中，平均值 $\mu\bar{x}$ 與原來全體的平均值 μ 是相等的。

② $\sigma\bar{x} = \frac{\sigma^2}{n} = \frac{\sigma}{\sqrt{n}}$，即樣本平均值 $\bar{x}$ 的分布中，變異數 $\sigma\bar{x}^2$ 與原來全體的變異數 σ^2 除以樣本數相等。

③ $\mu s^2 = \sigma^2$，樣本變異數的平均 μs^2 與全體變異數 σ^2 相等。

由於選出之樣本，往往並未放回原處，因此，$\sigma\bar{x}^2$ 的數值應校正如

下：

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

n：樣本數

N：全體數

又由於所得知的全體數，n 比 N 小得多，可省略不計，因此，$\sigma_{\bar{x}}^2 = \sigma^2/n$ 依舊可用。

差異的抽樣誤差（The Sampling Error of A Pifference）

在研究統計的資料時，由於須討論到二數量間的差異問題，因此，便涉及到二數量間差異的抽樣誤差問題。

設 y_1 與 y_2 爲二隨機變數，於 y_1 的分布中，m_1 爲平均值，v_1 爲其變異數，於 y_2 的分布中，m_2 爲平均值，v_2 爲其變異數，若分別自 y_1 與 y_2 中隨意抽取不同的一個觀察值，在重複若干次的抽樣後，$y_1 - y_2$ 的分布情形是怎樣的？

$$E(y_1 - y_2) = E(y_1) - E(y_2) = m_1 - m_2$$

$$\begin{aligned}
\mathrm{var}(y_1 - y_2) &= E\left[(y_1 - y_2) - (m_1 - m_2)\right]^2 \\
&= E\left[(y_1 - m_1) - (y_2 - m_2)\right]^2 \\
&= E\left[(y_1 - m_1)^2 - 2E(y_1 - m_1)\right. \\
&\quad \left.(y_2 - m_2) + E(y_2 - m_2)^2\right] \\
&= v_1 + v_2
\end{aligned}$$

若

$$y_1 = \bar{x}_1,\ y_2 = \bar{x}_2$$

$$m_1 = \mu_1,\ v_1 = \sigma_1^2/n_1$$

$$m_2 = \mu_2,\ v_2 = \sigma_2^2/n_2$$

由上式得

$$E(\bar{x}-\bar{x}_2)=\mu_1-\mu_2$$

$$\operatorname{var}(\bar{x}_1-\bar{x}_2)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

同理，若 $y_1=P_1$，於樣本數 n_1 時，某一全體中之某種情況的發生率爲 π_1，又 $y_2=P_2$，在另一樣本數 n_2，某種情況的發生機率爲 π_2，則

$$n_1=\pi_1，v_1=\pi_1(1-\pi_1)/n_1；$$

$$n_2=\pi_2，v_2=\pi_2(1-\pi_2)/n_2；$$

則

$$E(P_1-P_2)=\pi_1-\pi_2$$

$$\operatorname{var}(P_1-P_2)=\frac{\pi_1(1-\pi_1)}{\pi_1}+\frac{\pi_2(1-\pi_2)}{\pi_2}$$

比較二變異數之估計值時，若樣本估計值爲 S_1^2 與 S_2^2，則

$$E(S_1^2-S_2^2)=\sigma_1^2-\sigma_2^2$$

$$\operatorname{var}(S_1^2-S_2^2)=\frac{2\sigma_1^4}{n_1-1}+\frac{2\sigma_2^4}{n_2-11}$$

變異數的抽樣誤差（The sampling Error of A Variance）

設 μ 爲一量性隨機變數 x 分布中的平均，σ^2 爲變異數，樣本大小爲 n 時，則變異數估計值（estimated variance）爲：

$$S^2=\frac{\Sigma(x_i-\bar{x})^2}{n-1}$$

在此分布中，重複若干次隨機抽樣，S^2 的數值在每個樣本之中互不相同，這還是屬於一隨機變數。S^2 的性質敘述如下：

S^2 的期望值

$$E(S^2)=\frac{1}{n-1}E\left[\Sigma(x_i-\bar{x})^2\right]$$

$$=\frac{1}{n-1}E\{\Sigma[(x_i-\mu)-(\bar{x}-\mu)]^2\}$$

$$=\frac{1}{n-1}E\{\Sigma(x_i-\mu)^2-2\Sigma(x_i-\mu)(\bar{x}-\mu)+\Sigma(\bar{x}-\mu)^2\}$$

$$=\frac{1}{n-1}E\{(x_i-\mu)^2-n(\bar{x}-\mu)^2\}$$

$$\because \Sigma(xi-\mu)(\bar{x}-\mu)=\Sigma(\bar{x}-\mu)^2=n(\bar{x}-\mu)^2$$

依定義　$E(xi-\mu)^2=\sigma^2$

$$E(\bar{x}-\mu)^2=var(\bar{x})=\sigma^2/n$$

$$\therefore E(S^2)=\frac{1}{n-1}\{n\sigma^2-n(\frac{\sigma^2}{n})\}$$

$$=\frac{1}{n-1}(n-1)\sigma^2$$

$$=\sigma^2$$

按上式，可說 S^2 爲 σ^2 之不偏估計值（unbiased estimate）。因此，全體變異數的理想估計值爲 s^2，以 $n-1$ 爲被除數，若以 n 爲分母，則期望值便爲 $(n-1)\sigma^2/n$，數值便小於 σ^2，由此可知，E(S) 並不等於 σ，且小於 σ。

S^2 的抽樣分布。設 x 的分布爲極常態分布，在此情況下 S^2 的分布是十分近似 x^2 分布（chisquare or chisquared），現在先來討論 x^2 的分布，再討論 S^2 的分布。

假設 X_1 爲變數 x的標準化偏量（standardized deviate），則

$$X_1=(x-\mu)/\sigma$$

X_1^2 是隨機變數，其數值爲正值，如此，X_1^2 的分布就叫做見有一自由度(one degree of freedom , IDF) 的 x^2 分布，或爲 $x_{(1)}^2$ 分布。有兩點須特加注意：

① $E(X_1^2) = E(x-\mu)^2/\sigma^2$

$= \sigma^2/\sigma^2$

$= 1$，即分布的平均值爲 1

② 可由常態分布中求百分位値

設 X_1 與 X_2 爲 X 的二獨立觀察值

$$\text{若}\quad X_2^2 = \frac{(xi-\mu)^2}{\sigma^2} + \frac{(x_2-\mu)^2}{\sigma^2}$$

X_2^2 便爲具有二自由度的 x^2 分布（ $X_{(2)}^2$ ）。X_2^2 的數值與 X_1^2 同爲正值。X_2^2 爲 X_1^2 的二獨立觀察值的和，即：

$$E(X_2^2) = 2E(X_1^2)$$

$$= 2$$

同理，若 xi 變數有幾個數，則

$$x_n^2 = \sum_{i=1}^{n} \frac{(xi-\mu)^2}{\sigma^2}$$

$$= \frac{\Sigma(xi-\mu)^2}{\sigma^2}$$

由上式可知 $x_{(n)}^2$ 是具有 n 個自由度的 x^2 分布，而 $E(x_n^2) = n$。當自由度增加時，x^2 分布也就愈近於對稱。由於它是幾個不同 $x_{(1)}^2$ 變數的總和，中心極限定理(central limit theorem)可証實 $x_{(n)}^2$ 於 n 增加時，它的數值越接近單位值，所以，$x_{(n)}^2$ 分布的變異數爲 2n。

x_n^2 的式子中，可知變數和全體平均 μ 差異平方值之總和，在 S^2 的式子中，則顯示變數及樣本平均 $\overline{x}$ 差異平方值的總和，已知

$$\Sigma(xi-\overline{x})^2 \leq (xi-\mu)^2$$

實際上$\Sigma(xi-\overline{x})^2/\sigma^2$的形態是爲$x^2_{(n-1)}$分布

$$S^2=\frac{\Sigma(xi-\overline{x})^2}{n-1}$$

$$=\frac{\sigma}{n-1}\times\frac{\Sigma(xi-\overline{x})^2}{\sigma^2}$$

$$=\frac{\sigma^2}{n-1}x^2_{(n-1)}$$

$$S^2=\frac{\sigma^2}{n-1}\text{乘上}x^2_{(n-1)}$$

故

$$E(S^2)=\frac{\sigma^2}{n-1}E(x^2_{(n-1)})$$

$$=\frac{\sigma^2}{n-1}(n-1)$$

$$=\sigma^2$$

又

$$var(S^2)=\frac{\sigma^4}{(n-1)^2}var(x^2_{(n-1)})$$

$$=\frac{\sigma^4}{(n-1)^2}2(n-1)$$

$$=\frac{2\sigma^4}{(n-1)}$$

上式的條件爲：必須由常態分布中抽樣。

百分率的抽樣誤差 The sampling Error of A Proportion

設 A 及 B 爲一無限大之全體中的二個個體，π 及 $1-\pi$ 分別爲其機

率，A類的個體數爲r，其樣本大小n的隨機樣本是呈二項分布，以下將解釋r的平均值及變異數之公式。

若一量性變數x中，每個A的數值是1，每個B的數值是0，亦即x是每個個體的數值，樣本數n中有r個A與n－r個B。

$$\begin{aligned}\Sigma x &= (r\times 1)+\{(n-r)\times 0\}\\ &= r\end{aligned}$$

又
$$\begin{aligned}\bar{x} &= \frac{r}{n}\\ &= P\end{aligned}$$

我們可視樣本百分率P爲x的樣本平均，至於有關P的抽樣差異性爲何？須先x的全體平均與其標準偏差，因爲

$$\begin{aligned}E(x) &= (\pi\times 1)+[(1-\pi)\times 0]\\ &= \pi\end{aligned}$$

又
$$\begin{aligned}\mathrm{var}(x) &= E(x^2)-[E(x)]^2\\ &= (\pi\times 1^2)+[(1-\pi)\times 0^2]-\pi^2\\ &= \pi(1-\pi)\end{aligned}$$

故
$$\mathrm{var}(\bar{x})=\frac{\pi(1-\pi)}{n}$$

以P代入上二式之$\bar{x}$，則

$$E(P)=\pi，\mathrm{var}(p)=\frac{\pi(1-\pi)}{n}$$

又
$$r=nP$$

$$E(r)=n\pi，\mathrm{var}(r)=n\pi(1-\pi)$$

當n接近無限大時，$\bar{x}$即（P）的分布近似常態分布。

二、資料的搜集及抽樣的定義

所謂抽樣調查（ sampling survey ）亦可簡稱爲抽查，乃是在一龐大的全體中，隨機抽取一部分爲樣本，然後進行研究與調查，此乃因爲全體數過於龐大，不可能一一加以分析研討，爲了省時、省力與省經費，抽查是很好的調查方法。

於抽樣調查時，須注意外在效度（ external validity ），也就是「樣本足以代表全體之程度」，若能夠完全代表，則表示抽取之樣本完全與全體之內函一致，如此即達 100％的外在效度。若完全不能代表，即表示樣本與全體完全不一樣，如此，外在效度便爲 0 。愈能代表全體之樣本，外在度愈高，愈不能代表全體之樣本，外在度愈低。

調查及登記是搜集資料的二種方法，這是最基本，也是最直接的搜集法，但此二者的含意却不盡相同。舉例言之，生物醫藥的研究，大致即爲調查，但若是自某一現象中分析在某一時間內的變化情形，則爲登記。

調查的方法又可分爲普查（ Census ）及抽樣調查（ Sampling survey ）二種。

①普查——又叫做全體調查（ Complete enumeration ），係對一個大全體進行全面的調查，常聽到的有人口普查、農業普查與工商普查等。

由於普查的調查面太廣泛，以研究對象的每一分子進行全盤的調查，於時間、精力與財物方面消耗甚多，因此，除了較重要，且爲必須的研究對象外，一般很少應用到普查。

②抽樣調查——亦叫做抽查，也就是在一全體中，隨機抽取一部分爲樣本，代表全體進行研究調查。

此抽樣調查的優點是省時、省力又省錢，一般在研究某事物時都是採用此種方法。若抽樣得法、分析正確，一樣可達相當的準確程度。

抽樣調查難免與實際情況有所偏差，這種偏差就叫做抽樣誤差（Sampling error），要減少抽樣誤差的程度，最好是增加樣本數，或是增加調查的次數，但其限度須以人力、時間及經費作基準，否則，抽樣調查無異於普查，失去它的意義。

依研究對象性質之不同，須選擇適當的抽樣方法，主要有二種：

①機率抽樣法（Probability sampling）

此機率抽樣法乃係隨機抽樣，全體中的每一分子被抽取爲樣本的機會均一致，並且要知道被抽選的機率爲何，方能以機率原理計算出其抽樣誤差。若抽樣準確度已定，便可擇一方便調查研究且省時省力的抽樣方法。若抽樣的經費已定，則可選擇準確度最高，誤差最小的抽樣方法，這是現代抽樣調查理論與技術所追求的目標。

以下介紹五種常用的機率抽樣方法：

①單純隨機抽樣法（simple random sampling）

這種單純隨機抽樣法是一種最簡單的方法，一全體中的每一分子被選中的機會均相同。使用此法所抽取的樣本就叫做隨機樣本，其出現的機率，可應用或然率來計算出來。

②分層抽樣法（stratified sampling）

在使用此法之前，必須先將性質相同的樣本合爲一組，即爲「層」（strata or subgroup）。當進行抽樣時，便自每一層中，用隨機抽樣法抽出幾個所需的樣本數。每一層中，被抽取的機會也是相同的，但由於各層中的組成單位數目並不一樣，故各層之機會不會相同。

自各層中抽取樣本，可根據一定的抽樣率（sampling ratio）來做，假設各層中的小全體各爲 N_1，N_2，N_3，……，N_k，各層中的抽樣大小各爲 n_1，n_2，n_3，……，n_k，抽樣率則爲

$$\frac{n}{N} = f = \text{常數}$$

即

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \cdots\cdots \frac{n_k}{N_k} = f$$

此抽樣法稱爲比例抽樣法(Proportionate sampling)。

③集落抽樣法(Cluster sampling)

進行此集落抽樣之先，須將全體依某種標準，分爲若干組(groups)，這些組也就構成樣本，即爲集落(Cluster)，由集落之中抽取樣本的方法就是集落抽樣法。

在一研究之全體中，隨機抽樣成集落後，以集落中的全部樣本爲研究對象，則爲一段法。此外，亦可自其中再抽選樣本以爲調查對象，稱爲二段法。

一般的都市、鄉村、城鎮所形成的集落，大多是自然形成的，依此地理區域特性集合而抽樣的方法可稱爲地域抽樣法(area sampling)

若集落之大小相同，於抽樣上便要方便多了，但，如果其大小不同，爲了求得高度的準確性，抽選的方法應依集落大小的比例爲標準，叫做機率比例抽樣法(sampling with probability proportionate to size)。

④多段抽樣法(Multi－stage sampling)

依照集落抽樣法，抽取出集落後，再自各集落中捕出意欲研究調查的對象，叫做二段抽樣法(two－stage sampling)，亦爲副次抽樣法(sub－sampling)。

若抽樣的步驟分爲三段，即爲三段抽樣法，依此類推。一般通稱爲多段抽樣法。

⑤系統抽樣法(Systematic sampling)

使用此系統抽樣法時，必須知道全體的數量，需要的樣本數是多少，先將抽樣的間隔決定後，再開始抽樣的工作。

這是當樣本數太大，隨機抽樣法不實用時，即可換用之方法。說明如下：

設N爲全體總數，n爲樣本數，$k=\frac{N}{n}$爲抽樣間隔，於1至k中任選一數r，則r，r+k，r+2k，r+3k，……，r+(n-1)k等n個數爲需要的樣本。例如，N爲15，n爲3，則

$$k=\frac{N}{n}=\frac{15}{3}=5$$

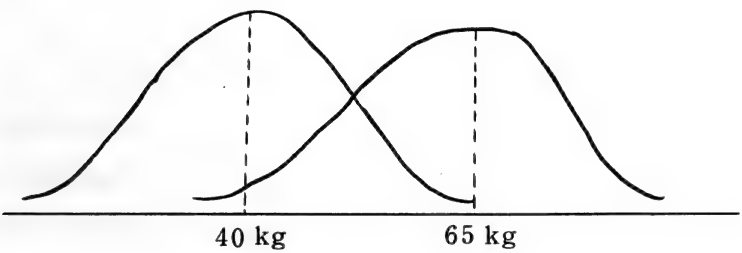
即每5個數選出一樣本。

系統抽樣亦有人使用雙向式抽樣法（two-dimentional sampling），也就是於横向及縱向均隔一定間隔來抽樣，可使抽樣的手續簡化些。

在進行系統抽樣法時，應注意在樣本範圍（Sampling frame）內的各個單位的排列，其與間隔是否有正相關的關係。若無此關係，則計算出的抽樣誤差會很大，因此，若是地區性的抽樣調查，應用系統抽樣時，要選擇不同的抽樣起點。

第十三章
常態分配的定義及性質

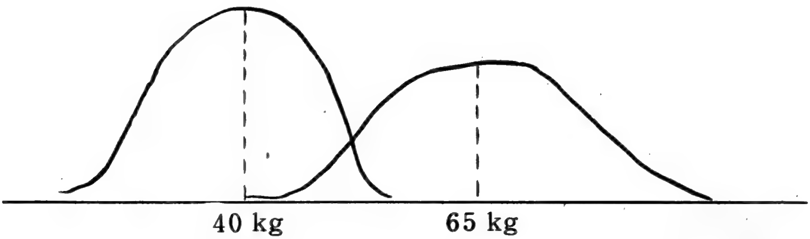
常態分配具有如下性質：常態次數函數與橫軸間的面積等於一個平方單位（ one square unit ）。依相對次數所製成的直方圖，該圖之面積必等於一，所以，常態曲線下的面積必也等於一，這是由於常態曲線（ normal curve ）是用來模擬直方圖的。常態曲線並不是只有一條，而是有許多條，只要有一個不同的標準差及均數，便會形成一不同之常態曲線。例如，成人的體重，可能十分近似常態分配，均數爲 公斤。十二歲的兒童的體重亦近似常態分配，均數爲 40 公斤，假使二者的體重標準差相等，那次數函數的形態也相同，只是另一線向右移動325公斤（見下圖）。



圖：均數不等，標準差相等之常態曲線

成人體重之差異很可能大於男童體重的差異。假設成人體重的標準差是5公斤，十二歲的兒童，其體重的標準差是3公斤，平均成人體重之距65公斤者要比兒童體重之距40公斤者較遠。所以，成人體重的次

數曲線要比兒童體重的次數曲線要平坦些，但亦較寬濶些，如下圖所示：



圖：標準差與均數皆不等的常態曲線

由上述可知，只要有一對 μ 與 σ，便能夠繪製一條常態次數曲線，且每一特定的 μ 及 σ 僅可繪製一條常態次數曲線。所以，假使已知體重成常態分配，均數爲 65 公斤，標準差爲 5 公斤，即可繪出正確之常態次數分配，且整個分配亦可繪成曲線圖。

上述二組資料，可能會有均數及標準差皆相同者，但也有次數完全不同的情形，所以，單是知道一組資料的均數及標準差，並不能夠繪出次數分配的情形。但，假如標準差及均數已知，並知其成常態分配，則整個次數便可完成一曲線圖。

一個呈常態分配的群體，標準差及均數可解釋它的全部。

依任二數 x_1 與 x_2 之間的直方圖之面積可決定此二數間的觀測值所占的比例。假使有一呈常態分配的群體，它介於 x_1 與 x_2 間的觀察值所占的比例等於在橫軸之上、常態曲線之下，也就是介於 x_1 與 x_2 所占的面積。

設 x_s 爲常態分配的觀測值，其均數爲 μ，標準差爲 σ，則若每一觀測值各減去 μ，所成之（ $x-\mu's$ ）新群體仍爲常態分配，該均數爲 0，標準差爲 σ 。之後，若將各（ $x-\mu$ ）分別除以 σ，（ $x-\mu$ ）/ σ 的群體也會呈常態分配，其均數爲 0，標準差爲 1 。如此，若作此 $z=$

($x-\mu$) / σ 如此之變換，所得到的群體z〔 z爲新變數(variable)稱爲標準常態變量(standard normal variate)〕亦當呈常態分配，該均數爲0，標準差爲1。

常態分配(normal distribution) 的重要性之一是很多大組的資料，均可由常態曲線密切模擬，於是稱此資料呈常態分配。此外，我們亦知，許多大組的資料並不呈常態分配，例如各個家庭的收入或是死亡的年齡都不可能完全呈常態分配。

若自任一群體中，取出一大樣本，之後再將與此等大的可能樣本全部取出，然後分別計算出該樣本的均數，則以這些均數所形成的群體勢必十分接近常態分配。不論群體之分配爲何，其樣本均數皆近似常態分配。

由樣本數所形成的群體均數必等於原來群體的均數，($\mu\overline{x}=\mu$)；而它的標準誤差也等於原來群體的標準差，被除以樣本大小的平方根($\sigma\overline{x}=\sigma/\sqrt{n}$)。只要樣本大小合理，樣本平均數會很近似常態分配。如果群體的μ及σ已知，則只要樣本大小合理，如此，樣本均數的分配之各方面，也就可得知了，可以不需要群體的分配是呈何形狀。

決定樣本數的大小，並不一定標準，可依二種情形而定：

①我們所期望的樣本均數的分配，其接近常態分配的程度爲何？

②原來群體的分配狀況爲何？

如果原來群體的分配與常態分配相差頗巨，便要比原來群體極接近常態分配所需的樣本數大。但一般的研究，樣本只要在 25 以上，假使樣本均數呈常態分配，誤差便不會太大，而原來群體的分配爲何，已無關重要，可不必考慮之。

第十四章
生物及醫學的實驗統計法則

一、中值檢定法的目的

若所研究對象的樣本資料不成對，則可使用中值檢定法(median test)，中值檢定法的目的和符號檢定法一樣。

$$T = \frac{\frac{n_1'}{n_1} - \frac{n_2'}{n_2}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

公式〔 14 － 1 〕

$$P = \frac{n_1' + n_2'}{n_1 + n_2}$$

$$q = 1 - P$$

此法是把樣本X及樣本Y的各變量由小至大編排，n_1 是X變量之個數，n_2 是Y變量的個數，n_1' 是X變量大於混合中位數的個數，n_2' 是Y變量大於混合中位數的個數。

原來全體平均數假設的考驗

平均數抽樣分配的標準記：

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

或：

$$\sigma_M = \frac{S}{\sqrt{N}}$$

公式〔 14 − 2 〕

或

$$\sigma_M = \frac{S}{\sqrt{N-1}}$$

$$T = \frac{M-\mu}{\sigma_M}$$

公式〔 14 − 3 〕

若爲有限母體（原來全體），在$\frac{N}{(N)} \leq \frac{1}{10}$時，可視爲無限群體，該$\sigma_M = \frac{S}{\sqrt{N}}$，或$\sigma_M = \frac{S}{\sqrt{N-1}}$；又在$\frac{N}{(N)} > \frac{1}{10}$時，便要視爲有限群體，該$\sigma_M = \frac{S}{\sqrt{N}}\sqrt{\frac{(N)-N}{(N)-1}}$，或$\sigma_M = \frac{S}{\sqrt{N-1}}\sqrt{\frac{(N)-N}{(N)-1}}$。總而言之，當爲有限母體時，若樣本次數及母體次數相差甚大時，便假設母體爲無限的。（N 爲樣本次數，（N）爲有限母體次數）

U 檢定法（The U test）

U 檢定法的作用與中值檢定法及符號檢定法一樣，但功用稍二者大。

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \Sigma〔n_1〕$$

公式〔 14 − 4 〕

或

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \Sigma〔n_2〕$$

公式〔 14 − 5 〕

n_1 ：X 樣本之個數

n_2 ：Y樣本之個數

$\Sigma[n_1]$：混合編排後（變量由小至大）的X樣本變量順序數之總和

$\Sigma[n_2]$：Y樣本變量順序數的總和

$$E(U)=\frac{n_1 n_2}{2}$$

公式〔14－6〕

$$\sigma U^2=\frac{n_1 n_2(n_1+n_2+1)}{12}$$

公式〔14－7〕

$$T=\frac{U-E(U)}{\sigma U}$$

公式〔14－8〕

若U及E（U）愈接近，則越近似兩母群體。若U與E（U）相差愈遠，則兩母群體愈不相近。

U檢定法還可應用下述公式〔14－9〕來計算T：

$$T=\frac{\Sigma[n_1]-E(\Sigma[n_1])}{\sigma U}$$

或：

$$T=\frac{\Sigma[n_2]-E(\Sigma[n_2])}{\sigma U}$$

公式〔14－9〕

$E(\Sigma[n_1])$與$E(\Sigma[n_2])$爲二個樣本等級數總和之期望值。

$$E(\Sigma[n_1])=\frac{n_1(n_1+n_2+1)}{2}$$

$$E(\Sigma[n_2]) = \frac{n_2(n_1 + n_2 + 1)}{2}$$

公式〔14－10〕

母體平均數之估計

95％的可信界限：

高限⇒$M + 1.96\,\sigma_M$

低限⇒$M - 1.96\,\sigma_M$

公式〔14－11〕

99％的可信界限：

高限⇒$M + 2.58\,\sigma_M$

低限⇒$M - 2.58\,\sigma_M$

公式〔14－12〕

二平均數之差的顯著性考驗與差數估計

$$\sigma M_1 - M_2 = \sqrt{\sigma M_1^2 + \sigma M_2^2}$$

$$= \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$$

公式〔14－13〕

或：

$$\sigma M_1 - M_2 = \sqrt{\frac{S_1^2}{N_1 - 1} + \frac{S_2^2}{N_2 - 1}}$$

公式〔14－14〕

當二母體的（N）不為無限大時，可用 $S_1^2\left(\frac{(N_1) - N_1}{(N_1) - 1}\right)$ 代替 S_1^2，而以 $S_2^2\left(\frac{(N_2) - N_2}{(N_2) - 1}\right)$ 代替 S_2^2。（N）為有限母體次數，N 不加括者為樣本次數。

$$T=\frac{(M_1-M_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_{M_1}{}^2+\sigma_{M_2}{}^2}}$$

公式〔14－15〕

平均數的差數估計

95％的可信界限：

$$(M_1-M_2)\pm 1.96\,\sigma_{M_1-M_2}$$

公式〔14－16〕

99％的可信界限：

$$(M_1-M_2)\pm 2.58\,\sigma_{M_1-M_2}$$

公式〔14－17〕

母體標準差的假設考驗與估計

$$\sigma_s=\frac{\sigma}{\sqrt{2N}}$$

$$\sigma_s=\frac{S}{\sqrt{2N}}$$

或：

$$\sigma_s=\frac{S}{\sqrt{2(N-1)}}$$

公式〔14－18〕

以95％的界限估計母體標準差範圍：

高限⇒ $S+1.96\,\sigma_s=S+1.96\dfrac{S}{\sqrt{2N}}$

低限⇒ $S+1.96\,\sigma_s=S+1.96\dfrac{S}{\sqrt{2N}}$

公式〔14－19〕

母體標準差是否相似的假設考驗

$$T=\frac{S_1-S_2}{\sigma_{s_1-s_2}}$$

$$\sigma_{s_1-s_2}=\sqrt{\sigma^2 s_1+\sigma^2 s_2}$$

公式〔14－20〕

母體相關係數假設之考驗和估計

由常態母體中二變量計算出之積差相關係數ρ，依統計法則，其可能各樣本相關係數r所形成的抽樣分配之形態，依樣本大小（N）與母全體常數（ρ）值的高低而變化，在$\rho=0$時，不管樣本大小其r的抽樣分配皆為對稱分配，若$\rho \neq 0$，其絕對值愈大或N愈小時，該抽樣分配愈偏斜。當ρ絕對值逐漸小至0.5或更小時，或大樣本時，r之抽樣分配可近似常態分配，則此抽樣分配之平均數與ρ近似，其標準差的近似值如下：

$$\sigma_r=\frac{1-\rho^2}{\sqrt{N-1}}$$

公式〔14－21〕

若以r代ρ，則

$$\sigma_r=\frac{1-r^2}{\sqrt{N-1}}$$

公式〔14－22〕

若二變量無關係，則$\rho=0$，但r則並不一定均為0，而以0為平均數，該標準誤如下：

$$\sigma_r=\frac{1-\sigma^2}{\sqrt{N-1}}$$

$$\sigma_r = \frac{1-0}{\sqrt{N-1}}$$

$$\sigma_r = \frac{1}{\sqrt{N-1}}$$

公式〔 14 － 23 〕

設 $\rho = 0$ ，r 爲大樣本時所得之相關，若要考驗母體相關是否爲0，也就是 H_0 ： $\rho = 0$ 能否成立，可由T值依1.96及2.58等顯著點決定原虛無假設母全體相關爲0應接受或棄之。

$$T = \frac{r-\rho}{\frac{1}{\sqrt{N-1}}}$$

$$= \frac{r-0}{\frac{1}{\sqrt{N-1}}}$$

公式〔 14 － 24 〕

若T值大於1.96，並大於2.56，不論依5％或1％水準，母體相關爲0的假設皆應棄之。

因樣本相關係數 r 的抽樣分配形態很不穩定，只在 ρ 的絕對值甚小而樣本次數很多時方能接近常態分配，所以，不應任意使用常態法則處理。

統計學家費宣氏(R. A. Fisher) 把 r 以簡單對數轉換成 Z'，把 ρ 轉換成 Z_0'，便不爲樣本大小與 ρ 值高低之左右，Z' 的抽樣分配均近似常態分配形態，可使用常態機率推斷。

$$Z' = 1.1513 \log_{10}\frac{1+r}{1-r}$$

$$Z_0' = 1.1513 \log_{10}\frac{1+\rho}{1-\rho}$$

公式〔 14 － 25 〕

計算時可參考下列二表：

自然對數表

N	$\log_e N$	N	$\log_e N$	N	$\log_e N$	N	$\log_e N$	N	$\log_e N$
0.1	7.6974	2.7	0.9933	5.3	1.6677	7.9	2.0669	15	2.7081
0.2	8.3906	2.8	1.0296	5.4	1.6864	8.0	2.0794	16	2.7726
0.3	8.7960	2.9	1.0647	5.5	1.7047	8.1	2.0919	17	2.8332
0.4	9.0837	3.0	1.0986	5.6	1.7228	8.2	2.1041	18	2.8904
0.5	9.3069	3.1	1.1314	5.7	1.7405	8.3	2.1163	19	2.9444
0.6	9.4892	3.2	1.1632	5.8	1.7579	8.4	2.1282	20	2.9957
0.7	9.6433	3.3	1.1939	5.9	1.7750	8.5	2.1401	25	3.2189
0.8	9.7769	3.4	1.2238	6.0	1.7918	8.6	2.1518	30	3.4012
0.9	9.8946	3.5	1.2528	6.1	1.8083	8.7	2.1633	35	3.5553
1.0	0.0000	3.6	1.2809	6.2	1.8245	8.8	2.1748	40	3.6889
1.1	0.0953	3.7	1.3083	6.3	1.8405	8.9	2.1861	45	3.8067
1.2	0.1823	3.8	1.3350	6.4	1.8563	9.0	2.1972	50	3.9120
1.3	0.2624	3.9	1.3610	6.5	1.8718	9.1	2.2083	55	4.0073
1.4	0.3365	4.0	1.3863	6.6	1.8871	9.2	2.2192	60	4.0943
1.5	0.4055	4.1	1.4110	6.7	1.9021	9.3	2.2300	65	4.1744
1.6	0.4700	4.2	1.4351	6.8	1.9169	9.4	2.2407	70	4.2485
1.7	0.5306	4.3	1.4586	6.9	1.9315	9.5	2.2513	75	4.3175
1.8	0.5378	4.4	1.4816	7.0	1.9459	9.6	2.2618	80	4.3820
1.9	0.6419	4.5	1.5041	7.1	1.9601	9.7	2.2721	85	4.4427
2.0	0.6931	4.6	1.5261	7.2	1.9741	9.8	2.2824	90	4.4998
2.1	0.7419	4.7	1.5476	7.3	1.9879	9.9	2.2925	95	4.5539
2.2	0.7885	4.8	1.5686	7.4	2.0015	10	2.3026	100	4.6052
2.3	0.8329	4.9	1.5892	7.5	2.0149	11	2.3979		
2.4	0.8755	5.0	1.6094	7.6	2.0281	12	2.4849		
2.5	0.9163	5.1	1.6292	7.7	2.0412	13	2.5649		
2.6	0.9555	5.2	1.6487	7.8	2.0541	14	2.6391		

積差相關係數 γ 轉換 Z' 值表

γ	Z'	γ	Z'	γ	Z'	γ	Z'
0.00	0.00	0.25	0.26	0.50	0.55	0.75	0.97
0.01	0.01	0.26	0.27	0.51	0.56	0.76	1.00
0.02	0.02	0.27	0.28	0.52	0.58	0.77	1.02
0.03	0.03	0.28	0.29	0.53	0.59	0.78	1.05
0.04	0.04	0.29	0.30	0.54	0.60	0.79	1.07
0.05	0.05	0.30	0.31	0.55	0.62	0.80	1.10
0.06	0.06	0.31	0.32	0.56	0.63	0.81	1.13
0.07	0.07	0.32	0.33	0.57	0.65	0.82	1.16
0.08	0.08	0.33	0.34	0.58	0.66	0.83	1.19
0.09	0.09	0.34	0.35	0.59	0.68	0.84	1.22
0.10	0.10	0.35	0.37	0.60	0.69	0.85	1.26
0.11	0.11	0.36	0.38	0.61	0.71	0.86	1.29
0.12	0.12	0.37	0.39	0.62	0.73	0.87	1.33
0.13	0.13	0.38	0.40	0.63	0.74	0.88	1.38
0.14	0.14	0.39	0.41	0.64	0.76	0.89	1.42
0.15	0.15	0.40	0.42	0.65	0.78	0.90	1.47
0.16	0.16	0.41	0.44	0.66	0.79	0.91	1.53
0.17	0.17	0.42	0.45	0.67	0.81	0.92	1.59
0.18	0.18	0.43	0.46	0.68	0.83	0.93	1.66
0.19	0.19	0.44	0.47	0.69	0.85	0.94	1.74
0.20	0.20	0.45	0.48	0.70	0.87	0.95	1.83
0.21	0.21	0.46	0.50	0.71	0.89	0.96	1.95
0.22	0.22	0.47	0.51	0.72	0.91	0.97	2.09
0.23	0.23	0.48	0.52	0.73	0.93	0.98	2.30
0.24	0.24	0.49	0.54	0.74	0.95	0.99	2.65

觀察上述二公式〔14－24〕、〔14－25〕，可得知γ自－1到1時，Z'自$-\infty$到∞，二者的符號一樣，在$\gamma=0$時，Z'也為0，因樣本的Z'之分配近似常態分配，該平均數即為母全體之常數Z_0'，其標準誤如下：

$$\sigma_z' = \frac{1}{\sqrt{N-3}}$$

公式〔14－26〕

以Z'值測定母全體的$Z_0'=0$的虛無假說能否成立，可應用下列公式：

$$T = \frac{Z'-0}{\dfrac{1}{\sqrt{N-3}}}$$

公式〔14－27〕

若T值大於1.96與2.58，便知依5％及1％水準，皆應棄之此假說。

考驗H_0：$\rho=0.55$能否成立，已知$N=100$，$\gamma=0.602$，則可使用下式：

$$T = \frac{Z'-Z_0'}{\dfrac{1}{\sqrt{N-3}}}$$

公式〔14－28〕

查表得知$\gamma=0.602$，$Z'=0.694$，$\rho=0.55$，$Z_0'=0.62$

則

$$T = \frac{0.694-0.62}{\dfrac{1}{\sqrt{97}}}$$

$$= 0.73$$

T值0.73之值甚小，故$\rho=0.55$的假設可成立。

母全體二變量積差相關係數ρ可信界限的估計

這項估計必須先把樣本的γ轉換爲Z′，計算出Z_0'的可信範圍，最後再轉換爲原來ρ的信賴界限。

95%的信賴界限：

$$Z' \pm 1.96 \frac{1}{\sqrt{N-3}}$$

公式〔14－29〕

99%的信賴界限：

$$Z' \pm 2.58 \frac{1}{\sqrt{N-3}}$$

公式〔14－30〕

二相關係數γ_1及γ_2的差量顯著性的考驗差量之標準誤公式如下：

$$\sigma_{Z_1'-Z_2'} = \sqrt{\sigma^2_{Z_1'} + \sigma^2_{Z_2'}} = \sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}$$

公式〔14－31〕

$$T = \frac{(N_1' - Z_2') - (Z_{01}' - Z_{02}')}{\sigma_{Z_1'-Z_2'}}$$

$$= \frac{Z_1' - Z_2' - 0}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}}$$

公式〔14－32〕

二、常態分配及二項分配的應用

測定母全體比率的假設，依樣本次數或樣本比例皆可，可歸納如下二式：

$$T=\frac{|f-f_0|}{\sqrt{Np_0q_0}}$$

公式〔 14 － 33 〕

$$T=\frac{|p-p_0|}{\sqrt{\frac{p_0q_0}{N}}}$$

公式〔 14 － 34 〕

f ：樣本次數

f_0 ：假設的理論次數（ $f_0=Np_0$ ）

N ：樣本項數

p_0 ：假設的母全體比率

q_0 ： $q_0=1-p_0$

p ：樣本比率

若為二端考驗，上二式中之T值大於或等於 1.96 時，則假設被摒棄於5％的水準上，若T值大於或等於 2.58 時，在1％的水準下，假設被棄之。

$\sqrt{\frac{p_0q_0}{N}}$ 為樣本比率抽樣分配的標準誤，也就是比率的標準誤，一般是以符號 σ_p 表示。若母全體不太大時，$\sigma_p=\sqrt{\frac{p_0q_0}{N}}\sqrt{\frac{(N)-N}{(N)-1}}$ 〔（N）為母體次數，N為樣本次數〕。

母體比率的估計

95％的信賴界限：

$$p_{OL}=p-1.96\sqrt{\frac{pq}{N}}$$

$$p_{OU}=p+1.96\sqrt{\frac{pq}{N}}$$

公式〔 14 － 35 〕

99％的信賴界限：

$$p_{OL}=p-2.58\sqrt{\frac{pq}{N}}$$

$$p_{OU}=p+2.58\sqrt{\frac{pq}{N}}$$

公式〔14－36〕

若爲有限母體時，95％的範圍可改爲：

$$p\pm 1.96\sqrt{\frac{pq}{N}}\sqrt{\frac{(N)-N}{(N)-1}}$$

公式〔14－37〕

若爲有限母體時，99％的範圍則爲：

$$p\pm 2.58\sqrt{\frac{pq}{N}}\sqrt{\frac{(N)-N}{(N)-1}}$$

（N）爲有限母體次數

N 爲樣本次數

二樣本比率之差異的顯著性之考驗：

$$T=\frac{(p_1-p_2)(p_0'-p_0'')}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

$$=\frac{p_1-p_2-0}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

公式〔14－38〕

$$\bar{p}=\frac{N_1'+N_2}{N_1+N_2}=\frac{N_1P_1+N_2P_2}{N_1+N_2}$$

$$\bar{q}=1-\bar{P}$$

成對資料間的差異顯著性的符號考驗

檢定二樣本是否源自於同一母全體乃爲符號檢定法之目的。由於此法是以二樣本變數相差的正負符號的多少作爲檢定的標準，因此名爲符號檢定法。

$$T = \frac{\left(n' + \frac{1}{2}\right) - np_0}{\sqrt{np_0 q_0}}$$

公式〔14－39〕

$$T = \frac{\left(n' - \frac{1}{2}\right) - np_0}{\sqrt{np_0 q_0}}$$

公式〔14－40〕

n：樣本中不爲0之次數

n'：第一樣本減第二樣本的正號之次數

當$n' < np_0$時，應用公式〔14－39〕，若$n' > np_0$時，應用公式〔14－40〕。

$$p_0 = q_0 = \frac{1}{2}$$

三、二項分配之性質

二項分配原本係來自於品質資料的抽樣，而現逐漸演化爲依據樣本資料具有某一屬性的比率，用來考驗有關母全體該項屬性比率的某種假設，或者估計母全體具有同樣屬性的比率。此乃依機率原理及抽樣理論而成的抽樣分配。

設自52張的撲克牌中，欲抽取一張紅桃10，其出現紅桃10的機率爲1/52，不爲紅桃10的機率爲51/52。

假設投擲一枚硬幣，正面朝上的原來機率爲 p_0，反面朝上的原來機率爲 q_0，$p_0 + q_0 = 1$，則投擲硬幣一枚其正反二面之機率各爲1/2

，可表示如下：

$$(p_0+q_0)^2=p_0^2+2p_0q_0+q_0^2$$

$$\left(\frac{1}{2}+\frac{1}{2}\right)^2=\frac{1}{4}+\frac{1}{2}+\frac{1}{4}$$

$$(p_0+q_0)^3=p_0^3+3p_0^2q_0+3p_0q_0^2+q_0^3$$

$$\left(\frac{1}{2}+\frac{1}{2}\right)^3=\left(\frac{1}{2}\right)^3+3\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3$$

$$=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}$$

若投擲一枚硬幣若干次 n，其機率的分配如下：

$$(q_0+p_0)^n=q_0^n+nq_0^{n-1}p_0+\frac{n(n-1)}{2}q_0^{n-2}p_0^2\cdots+\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}q_0^{n-r}p_0^r+\cdots+nq_0p_0^{n-1}+p_0^n$$

依上式出現的機率或相對次數如下：

正面的個數	機　　率
0	q_0^n
1	$nq_0^{n-1}p_0$
2	$\frac{n(n-1)}{2!}q_0^{n-2}p_0^2$
⋮	⋮
r	$\frac{n(n-1)(n-2)\cdots(n-Y+1)}{r!}q_0^{n-r}p_0^r$
⋮	⋮
n－1	$nq_0p_0^{n-1}$
n	p_0^n
總計	1

二項分配的峯度係數、偏態係數、算術平均數及標準差，計算公式如下：

峯度係數

$$K = 3 + \frac{1 - 6p_0 q_0}{np_0 q_0}$$

公式〔14－41〕

偏態係數

$$S.K. = \frac{q_0 - p_0}{\sqrt{np_0 q_0}}$$

公式〔14－42〕

算術平均數

$$\mu = np_0$$

公式〔14－43〕

標準差

$$\sigma = \sqrt{np_0 q_0}$$

公式〔14－44〕

觀察以上四個公式，可知二項分配的μ（即中心位置）隨樣本項數n及正面出現的機率p_0而改變。若p_0爲固定，μ隨樣本項數增加或減少而變大或變小，它的離差也是同樣的。

若$p_0 = q_0 = \frac{1}{2}$，樣本項數不論爲何，S. K. 皆爲0且成對種分配。當$p_0 < \frac{1}{2}$，即$p_0 < q_0$時，S. K. 爲正，若$p_0 > \frac{1}{2}$，即$p_0 > q_0$時，S. K. 爲負。再者，若$p_0 \neq q_0 \neq \frac{1}{2}$時，若n甚大，S. K. 也可趨向近0而爲對稱分配。

若不考慮峯度係數，當n甚大時，K便趨向3而成常態分配，由此可知，二項分配之形態，在n爲無限大時，二項分配也就近似常態分配，可依常態分配法則來處理。但一般情況下，n不可能爲無限大，所以

常態曲線很難符合二項分配而無誤，因此，僅可有效地利用常態曲線來估計二項分配機率。

一般計算二項分配（ $q_0 + p_0$ ）^{n} 的某一機率時，若 n 乘上 p_0 或 q_0（選一較小者）大於5，便可應用常態曲線來估計二項機率。

若 n 乘上 q_0 或 p_0 等於或小於 5，當 n 大於10，而 S. K. 小於0.2時，也可以使用常態曲線來估計二項機率。

在統計上，二項分配機率已有表可查，只要知道 n、r 與 p_0，便可利用下表查出機率爲何。

二項分配機率表 $p_0(r) = \binom{n}{r} q_0^{n-r} p_0^r$

n	r \ p_0	0.01	0.05	0.10	0.20
2	0	0.9801	0.9025	0.8100	0.6400
	1	0.0198	0.0950	0.1800	0.3200
	2	0.0001	0.0025	0.0100	0.0400
3	0	0.9703	0.8574	0.7290	0.5120
	1	0.0294	0.1354	0.2430	0.3840
	2	0.0003	0.0071	0.0270	0.0960
	3	0.0000	0.0001	0.0010	0.0080
4	0	0.9606	0.8145	0.6561	0.4096
	1	0.0388	0.1715	0.2916	0.4086
	2	0.0006	0.0135	0.0486	0.1536
	3	0.0000	0.0005	0.0036	0.0256
	4	0.0000	0.0000	0.0001	0.0016
5	0	0.9510	0.7738	0.5905	0.3277
	1	0.0480	0.2036	0.3280	0.4096

	2	0.0010	0.0214	0.0729	0.2048
	3	0.0000	0.0011	0.0081	0.0512
	4	0.0000	0.0000	0.0004	0.0064
	5	0.0000	0.0000	0.0000	0.0003
6	0	0.9415	0.7351	0.5314	0.2621
	1	0.0571	0.2321	0.3543	0.3932
	2	0.0014	0.0305	0.0984	0.2458
	3	0.0000	0.0021	0.0146	0.0819
	4	0.0000	0.0001	0.0012	0.0154
	5	0.0000	0.0000	0.0001	0.0015
	6	0.0000	0.0000	0.0000	0.0001
7	0	0.9321	0.6983	0.4783	0.2097
	1	0.0659	0.2573	0.3720	0.3670
	2	0.0020	0.0406	0.1240	0.2573
	3	0.0000	0.0036	0.0230	0.1147
	4	0.0000	0.0002	0.0026	0.0287
	5	0.0000	0.0000	0.0002	0.0043
	6	0.0000	0.0000	0.0000	0.0004
	7	0.0000	0.0000	0.0000	0.0000
8	0	0.9227	0.6634	0.4305	0.1678
	1	0.0746	0.2793	0.3826	0.3355
	2	0.0026	0.0615	0.1488	0.2936
	3	0.0001	0.0054	0.0331	0.1468
	4	0.0000	0.0004	0.0046	0.0459
	5	0.0000	0.0000	0.0004	0.0092

	6	0.0000	0.0000	0.0000	0.0011
	7	0.0000	0.0000	0.0000	0.0001
	8	0.0000	0.0000	0.0000	0.0000
9	0	0.9135	0.6302	0.3874	0.1342
	1	0.0830	0.2985	0.3874	0.3020
	2	0.0034	0.0629	0.1722	0.3020
	3	0.0001	0.0077	0.0446	0.1762
	4	0.0000	0.0006	0.0074	0.0661
	5	0.0000	0.0000	0.0008	0.0165
	6	0.0000	0.0000	0.0001	0.0028
	7	0.0000	0.0000	0.0000	0.0003
	8	0.0000	0.0000	0.0000	0.0000
	9	0.0000	0.0000	0.0000	0.0000
10	0	0.9044	0.5987	0.3487	0.1074
	1	0.0914	0.3151	0.3874	0.2684
	2	0.0042	0.0746	0.1937	0.3020
	3	0.0001	0.0105	0.0574	0.2013
	4	0.0000	0.0010	0.0112	0.0881
	5	0.0000	0.0001	0.0015	0.0264
	6	0.0000	0.0000	0.0001	0.0055
	7	0.0000	0.0000	0.0000	0.0008
	8	0.0000	0.0000	0.0000	0.0001
	9	0.0000	0.0000	0.0000	0.0000
	10	0.0000	0.0000	0.0000	0.0000

續上表 p_0 爲0.20之後

n	p_0 / r	0.25	$\frac{1}{3}$	0.40	0.50
2	0	0.5625	0.4444	0.3600	0.2500
	1	0.3750	0.4444	0.4800	0.5000
	2	0.0625	0.1111	0.1600	0.2500
3	0	0.4219	0.2963	0.2160	0.1250
	1	0.4219	0.4444	0.4320	0.3750
	2	0.1406	0.2222	0.2880	0.3750
	3	0.0156	0.0370	0.0640	0.1250
4	0	0.3164	0.1975	0.1296	0.0625
	1	0.4219	0.3951	0.3456	0.2500
	2	0.2109	0.2963	0.3456	0.3750
	3	0.0469	0.0988	0.1536	0.2500
	4	0.0039	0.0123	0.0256	0.0625
5	0	0.2373	0.1317	0.0778	0.0312
	1	0.3955	0.3292	0.2592	0.1562
	2	0.2637	0.3292	0.3456	0.3125
	3	0.0879	0.1646	0.2304	0.3125
	4	0.0146	0.0412	0.0768	0.1562
	5	0.0010	0.0041	0.0102	0.0312
6	0	0.1780	0.0878	0.0467	0.0156
	1	0.3560	0.2634	0.1866	0.0938
	2	0.2966	0.3292	0.3110	0.2344
	3	0.1318	0.2195	0.2765	0.3125

	4	0.0330	0.0823	0.1382	0.2344
	5	0.0044	0.0165	0.0369	0.0938
	6	0.0002	0.0014	0.0041	0.0156
7	0	0.1335	0.0585	0.0280	0.0078
	1	0.3115	0.2048	0.1306	0.0547
	2	0.3115	0.3073	0.2613	0.1641
	3	0.1730	0.2561	0.2903	0.2734
	4	0.0577	0.1280	0.1935	0.2734
	5	0.0115	0.0384	0.0774	0.1641
	6	0.0013	0.0064	0.0172	0.0547
	7	0.0001	0.0005	0.0016	0.0078
8	0	0.1001	0.0390	0.0168	0.0039
	1	0.2670	0.1561	0.0896	0.0312
	2	0.3115	0.2731	0.2090	0.1094
	3	0.2076	0.2731	0.2787	0.2188
	4	0.0865	0.1707	0.2322	0.2734
	5	0.0231	0.0683	0.1239	0.2188
	6	0.0038	0.0171	0.0413	0.1094
	7	0.0004	0.0024	0.0079	0.0312
	8	0.0000	0.0002	0.0007	0.0039
9	0	0.0751	0.0260	0.0101	0.0020
	1	0.2253	0.1171	0.0605	0.0176
	2	0.3003	0.2341	0.1612	0.0703
	3	0.2336	0.2731	0.2508	0.1641
	4	0.1168	0.2048	0.2508	0.2461

	5	0.0389	0.1024	0.1672	0.2461
	6	0.0087	0.0341	0.0743	0.1641
	7	0.0012	0.0073	0.0212	0.0703
	8	0.0001	0.0009	0.0035	0.0170
	9	0.0000	0.0001	0.0003	0.0020
10	0	0.0563	0.0173	0.0060	0.0010
	1	0.1877	0.0867	0.0403	0.0098
	2	0.2818	0.1951	0.1209	0.0439
	3	0.2503	0.2601	0.2150	0.1172
	4	0.1460	0.2276	0.2508	0.2051
	5	0.0584	0.1366	0.2007	0.2461
	6	0.0162	0.0569	0.1115	0.2051
	7	0.0031	0.0163	0.0425	0.1172
	8	0.0004	0.0030	0.0106	0.0439
	9	0.0000	0.0003	0.0016	0.0098
	10	0.0000	0.0000	0.0001	0.0010

常態曲線的性質

①常態曲線之方程式

ⓐ常態機率曲線方程式

二項分配，若次數N愈大，也就愈接近常態分配。

$$P(x)=\frac{1}{\sqrt{2\pi NP_0q_0}}e^{-\frac{(X-NP_0)^2}{2NP_0q_0}}$$

$NP_0=\mu$ 爲平均數

$\sqrt{NP_0q_0}=\sigma$ 爲標準差

$$P(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

公式〔14－45〕

P（x）：變量x的相對次數，也就是絕對次數和總次數的比值，也是變量x的機率。

σ：標準差

μ：平均數

π：爲一常數（3.14159）

e：亦爲一常數，是自然對數之底（約爲2.71828）

ⓑ常態曲線的標準形式

$$f(w)=y=\frac{N}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{w^2}{2}}$$

公式〔14－46〕

由公式〔14－46〕可知常態曲線的標準形式，乃係設總面積爲一個單位，$\sigma=1$，$\mu=0$。公式〔14－46〕中的標準值即爲w值，$w=\frac{x-\mu}{\sigma}$，此外，$f(w)$爲各標準值w上的縱坐標的高，也就是相對次數。

觀察公式〔14－46〕，可知，對稱縱軸（y_0）其標準差爲1，平均數爲0的常態分配曲線也就是標準常態曲線方程式。

② 常態曲線的高度及面積

常態曲線的高度及面積可查常態分配機率表與下表常態曲線縱坐標表：

常態曲線縱坐標表

$\frac{x}{\sigma}$(或Z,W,T)	f(w)	$\frac{x}{\sigma}$	f(w)	$\frac{x}{\sigma}$	f(w)
0.00	0.3989423	0.06	0.3982248	0.12	0.3960802
0.01	0.3989223	0.07	0.3979661	0.13	0.3955854
0.02	0.3988625	0.08	0.3976677	0.14	0.3950517
0.03	0.3987628	0.09	0.3973298	0.15	0.3944793
0.04	0.3986233	0.10	0.3969525	0.16	0.3938684
0.05	0.3984439	0.11	0.3965360	0.17	0.3923190
0.18	0.3925315	0.54	0.3448180	0.90	0.2660852
0.19	0.3918060	0.55	0.3429439	0.91	0.2636880
0.20	0.3910427	0.56	0.3410548	0.92	0.2612863
0.21	0.3902419	0.57	0.3391243	0.93	0.2588805
0.22	0.3894038	0.58	0.3371799	0.94	0.2564713
0.23	0.3885286	0.59	0.3352132	0.95	0.2540591
0.24	0.3876166	0.60	0.3332246	0.96	0.2516443
0.25	0.3866681	0.61	0.3312147	0.97	0.2492277
0.26	0.3856834	0.62	0.3291840	0.98	0.2468095
0.27	0.3846627	0.63	0.3271330	0.99	0.2443904
0.28	0.3836063	0.64	0.3250625	1.00	0.2419707
0.29	0.3825146	0.65	0.3229724	1.01	0.2395511
0.30	0.3813878	0.66	0.3208638	1.02	0.2371320
0.31	0.3802264	0.67	0.3187371	1.03	0.2347138
0.32	0.3790305	0.68	0.3165929	1.04	0.2322970
0.33	0.3778007	0.69	0.3144317	1.05	0.2298821
0.34	0.3765372	0.70	0.3122539	1.06	0.2274696
0.35	0.3752403	0.71	0.3100603	1.07	0.2250599

（續上頁）

0.36	0.3739106	0.72	0.3078513	1.08	0.2226535
0.37	0.3725483	0.73	0.3056274	1.09	0.2202508
0.38	0.3711539	0.74	0.3033893	1.10	0.2178522
0.39	0.3697277	0.75	0.3011374	1.11	0.2154582
0.40	0.3682701	0.76	0.2988724	1.12	0.2130691
0.41	0.3667817	0.77	0.2965948	1.13	0.2106856
0.42	0.3652627	0.78	0.2943050	1.14	0.2083078
0.43	0.3637136	0.79	0.2920038	1.15	0.2059363
0.44	0.3621349	0.80	0.2896916	1.16	0.2035714
0.45	0.3605270	0.81	0.2873689	1.17	0.2012135
0.46	0.3588903	0.82	0.2850364	1.18	0.1988631
0.47	0.3572253	0.83	0.2826945	1.19	0.1965205
0.48	0.3555325	0.84	0.2803438	1.20	0.1941861
0.49	0.3538124	0.85	0.2779849	1.21	0.1918602
0.50	0.3520653	0.86	0.2756182	1.22	0.1895432
0.51	0.3502919	0.87	0.2732444	1.23	0.1872354
0.52	0.3484925	0.88	0.2708640	1.24	0.1849373
0.53	0.3466677	0.89	0.2684774	1.25	0.1826491
1.26	0.1803712	1.62	0.1074061	1.98	0.0561831
1.27	0.1781038	1.63	0.1056748	1.99	0.0550789
1.28	0.1758474	1.64	0.1039611	2.00	0.0539910
1.29	0.1736022	1.65	0.1022649	2.01	0.0529192
1.30	0.1713686	1.66	0.1005864	2.02	0.0518636
1.31	0.1691468	1.67	0.0989255	2.03	0.0508239
1.32	0.1669370	1.68	0.0972823	2.04	0.0498001

（續上頁）

1.33	0.1647397	1.69	0.0956568	2.05	0.0487920
1.34	0.1625551	1.70	0.0940491	2.06	0.0477996
1.35	0.1603838	1.71	0.0924591	2.07	0.0468226
1.36	0.1582248	1.72	0.0908870	2.08	0.0458611
1.37	0.1560797	1.73	0.0893326	2.09	0.0449148
1.38	0.1539483	1.74	0.0877961	2.10	0.0439836
1.39	0.1518308	1.75	0.0872663	2.11	0.0430674
1.40	0.1497275	1.76	0.0847764	2.12	0.0421661
1.41	0.1476385	1.77	0.0832932	2.13	0.0412795
1.42	0.1455641	1.78	0.0818278	2.14	0.0404076
1.43	0.1435046	1.79	0.0803801	2.15	0.0395500
1.44	0.1414600	1.80	0.0789502	2.16	0.0387069
1.45	0.1394306	1.81	0.0775379	2.17	0.0378779
1.46	0.1374165	1.82	0.0761433	2.18	0.0370626
1.47	0.1354181	1.83	0.0747653	2.19	0.0362619
1.48	0.1334358	1.84	0.0734068	2.20	0.0354746
1.49	0.1314684	1.85	0.0720649	2.21	0.0347009
1.50	0.1295176	1.86	0.0707404	2.22	0.0339408
1.51	0.1275830	1.87	0.0694333	2.23	0.0331939
1.52	0.1256646	1.88	0.0681436	2.24	0.0324603
1.53	0.1237628	1.89	0.0668711	2.25	0.0317397
1.54	0.1218775	1.90	0.0656168	2.26	0.0310319
1.55	0.1200090	1.91	0.0643877	2.27	0.0303370
1.56	0.1181573	1.92	0.0631566	2.28	0.0296546
1.57	0.1163225	1.93	0.0619524	2.29	0.0289347

（續上頁）

1.58	0.1145048	1.94	0.0607652	2.30	0.0283270
1.59	0.1127042	1.95	0.0595947	2.31	0.0276816
1.60	0.1109208	1.96	0.0584409	2.32	0.0270481
1.61	0.1091548	1.97	0.0573038	2.33	0.0264265
2.34	0.0258166	2.70	0.0104209	3.06	0.0036951
2.35	0.0252182	2.71	0.0101428	3.07	0.0035836
2.36	0.0246313	2.72	0.0098712	3.08	0.0034751
2.37	0.0240556	2.73	0.0096058	3.09	0.0033695
2.38	0.0234910	2.74	0.0093466	3.10	0.0032668
2.39	0.0229347	2.75	0.0090936	3.11	0.0031669
2.40	0.0223945	2.76	0.0088465	3.12	0.0030608
2.41	0.0218264	2.77	0.0086052	3.13	0.0029754
2.42	0.0213407	2.78	0.0083697	3.14	0.0028835
2.43	0.0208294	2.79	0.0081398	3.15	0.0027945
2.44	0.0203284	2.80	0.0079155	3.16	0.0027075
2.45	0.0198374	2.81	0.0076925	3.17	0.0026231
2.46	0.0193563	2.82	0.0074829	3.18	0.0025412
2.47	0.0188350	2.83	0.0072744	3.19	0.0024615
2.48	0.0184233	2.84	0.0070711	3.20	0.0023841
2.49	0.0179711	2.85	0.0068728	3.21	0.0023089
2.50	0.0175283	2.86	0.0066793	3.22	0.0022358
2.51	0.0170947	2.87	0.0064907	3.23	0.0021649
2.52	0.0166701	2.88	0.0063067	3.24	0.0020960
2.53	0.0162545	2.89	0.0061274	3.25	0.0020290
2.54	0.0158476	2.90	0.0059525	3.26	0.0019641

（續上頁）

2.55	0.0154493	2.91	0.0057821	3.27	0.0019010
2.56	0.0150596	2.92	0.0056160	3.28	0.0018397
2.57	0.0146782	2.93	0.0054541	3.29	0.0017803
2.58	0.0143051	2.94	0.0052963	3.30	0.0017226
2.59	0.0139401	2.95	0.0051426	3.31	0.0016666
2.60	0.0135830	2.96	0.0049929	3.32	0.0016122
2.61	0.0132387	2.97	0.0048470	3.33	0.0015595
2.62	0.0128921	2.98	0.0047050	3.34	0.0015084
2.63	0.0125581	2.99	0.0045666	3.35	0.0014587
2.64	0.0122315	3.00	0.0044318	3.36	0.0014106
2.65	0.0119122	3.01	0.0043007	3.37	0.0013639
2.66	0.0116001	3.02	0.0041729	3.38	0.0013187
2.67	0.0112951	3.03	0.0040486	3.39	0.0012748
2.68	0.0109969	3.04	0.0039276	3.40	0.0012322
2.69	0.0107056	3.05	0.0038093	3.41	0.0011910
3.42	0.0011510	3.64	0.0005294	3.86	0.0002320
3.43	0.0011122	3.65	0.0005105	3.87	0.0002232
3.44	0.0010747	3.66	0.0004921	3.88	0.0002147
3.45	0.0010383	3.67	0.0004744	3.89	0.0002065
3.46	0.0010030	3.68	0.0004573	3.90	0.0001987
3.47	0.0009689	3.69	0.0004408	3.91	0.0001910
3.48	0.0009358	3.70	0.0004248	3.92	0.0001837
3.49	0.0009037	3.71	0.0004093	3.93	0.0001766
3.50	0.0008727	3.72	0.0003944	3.94	0.0001698
3.51	0.0008426	3.73	0.0003800	3.95	0.0001633

（續上頁）

3.52	0.0008135	3.74	0.0003661	3.96	0.0001569
3.53	0.0007853	3.75	0.0003526	3.97	0.0001508
3.54	0.0007581	3.76	0.0003306	3.98	0.0001449
3.55	0.0007371	3.77	0.0003271	3.99	0.0001393
3.56	0.0007061	3.78	0.0003149	4.00	0.0001338
3.57	0.0006814	3.79	0.0003032	4.1	0.0000893
3.58	0.0006575	3.80	0.0002919	4.2	0.0000589
3.59	0.0006313	3.81	0.0002810	4.3	0.0000385
3.60	0.0006119	3.82	0.0002705	4.4	0.0000249
3.61	0.0005902	3.83	0.0002604	4.6	0.0000101
3.62	0.0005693	3.84	0.0002506	4.8	0.0000040
3.63	0.0005490	3.85	0.0002411	5.0	0.0000015

查閱常態分配機率表及常態曲線縱坐標表，不但可依 w 值某出曲線下某部分面積與某點之高度，尚可自曲線下某部分面積查出所占標準量 R 上的距離。

③常態分配的性質

ⓐ常態分配是屬於單峯對稱分配，在中間部分有一最高點，分別向左右二邊逐漸下降，而形成兩邊對稱的形狀。常態分配中的算術平均數、中位數與衆數的數值皆一致，皆位於基線中間的同一點上。以公式〔14－46〕來說，標準值 $w=0$ 的位置也就是算術平均數的位置。

由於常態分配對稱在縱軸 y_0 ，因而可知常態分配之算術平均數及中位數等值。再者，若 $w=0$ 的位置，它的縱坐標是最高的位置，若將 $w=0$ 代入公式〔14－46〕

$$y = f(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$$

得 $y_0 = f(0) = \frac{1}{\sqrt{2\pi}} = 0.39894$，因爲這是縱坐標最高的位置，也就是次數最密的一點，因此，在基線之點就是衆數，而算術平均數又與衆數同值。

以 $\frac{df(w)}{dw}$ 與 $\frac{d^2f(w)}{dw^2}$ 來說，若 $w=0$，則

$$\frac{df(w)}{dw} = -\frac{w}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} = 0$$

而：

$$\frac{d^2f(w)}{dw^2} = \frac{w^2}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} = -\text{號}$$

依微分法則 $\frac{df(w)}{dw} = 0$，$\frac{d^2f(w)}{dw^2}$ 爲負號時是最大的一點，也就是說 $f(w)$ 的值最大，次數也最密。公式〔14－46〕中，標準值 $w=0$ 時，可符合此法則，並且也可知常態分配的算術平均數和衆數是同值的。

ⓑ曲線彎曲方向的轉變點，就叫做反曲點，標準常態曲線的二個反曲點位於標準值 $w=\pm 1$ 的位置。就微分法則來說，當 $\frac{d^2f(w)}{dw^2}=0$ 時，便能知道反曲點的位置。現

$$\frac{d^2f(w)}{dw^2} = \frac{w}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$$

當 $w=\pm 1$ 時，$\frac{d^2f(w)}{dw^2}$ 爲 0，因此可知常態曲線的反曲點位於 $w=\pm 1$ 的位置。

ⓒ常態分配的圖形，按其標準差及平均數的不同有所改變，因此，這種圖形的種類甚多，一般常應用到的是公式〔14－46〕的標準常態曲線。常態分配的峯度係數爲3，偏度係數爲0，這是用來決定某次的分配是否近常態分配的根據。

ⓓ常態分配的數值有無限多，這是由於常態分配是一種連續分配的結果，常態分配的變量是介於－∞到∞之間，因此，常態曲線的二端漸近基線，但却永遠不會相交。

④次數分配表的常態性適合度之考驗

以樣本次數分配表來考驗母全體次數的常態性適合度的方法，一般應用最爲廣泛的有卡方考驗法及a（Alpha）考驗法二種，在此，暫不討論卡方考驗法。

a考驗法乃指依據樣本的偏度量數及峯度量數來運用的考驗。它的功用除了可考驗母全體是否接近常態分配以外，同時也能顯示它的非常態性質是因偏態或峯態或二者皆有關。

查閱下表可知，在一常態母體中取一N＝150之隨機樣本，90％的偏態係數在±0.32之間，98％的偏態係數位在0.46之間，90％的峯度係數是位於2.45～3.65之間，98％的峯度係數是位於2.26～4.14之間。在使用時，若樣本的偏度係數及峯度係數都在90％的界限內，母全體的常態性是可靠的，若有其中一數位於90％界限以外，98％界限之內時，該母全體的常態性便不可靠了。此外，若有其中任一數位於98％以外，該母全體的常態性便不易成立了。

圖爲常態母全體的偏度及峯度之抽樣界限表

樣本項數	a_3（偏度）		a_4（峯度）	
	90%	98%	90%	98%
50	±0.53	±0.79	—	—
100	±0.39	±0.57	2.35 – 3.77	2.18 – 4.39
125	±0.35	±0.51	2.40 – 3.70	2.24 – 4.24
150	±0.32	±0.46	2.45 – 3.65	2.26 – 4.14
175	±0.30	±0.43	2.48 – 3.61	2.33 – 4.05
200	±0.28	±0.40	2.51 – 3.57	2.37 – 3.98
250	±0.25	±0.36	2.55 – 3.52	2.42 – 3.87
300	±0.23	±0.33	2.59 – 3.47	2.46 – 3.79
350	±0.21	±0.30	2.62 – 3.44	2.50 – 3.72
400	±0.20	±0.28	2.64 – 3.41	2.52 – 3.67
500	±0.18	±0.26	2.67 – 3.37	2.57 – 3.60
600	±0.16	±0.23	2.70 – 3.34	2.60 – 3.54
700	±0.15	±0.22	2.72 – 3.31	2.62 – 3.50
800	±0.14	±0.20	2.74 – 3.29	2.65 – 3.46
900	±0.13	±0.19	2.75 – 3.28	2.66 – 3.43
1000	±0.13	±0.18	2.76 – 3.26	2.68 – 3.41
1200	±0.12	±0.16	2.78 – 3.24	2.71 – 3.37
1600	±0.10	±0.14	2.81 – 3.21	2.74 – 3.32
2000	±0.09	±0.13	2.83 – 3.18	2.77 – 3.28
5000	±0.06	±0.08	2.89 – 3.12	2.75 – 3.17

四、統計推論的理論係由抽樣分配而來

統計方法通常可分爲推論統計（ Statistical estimation ）與敍述統計二種。敍述統計是使用簡單的數字代表樣本的各項統計常數。統計推論乃係依據樣本資料來推論或估計母全體（ Parent population ）之各種特性，例如，以樣本的平均數、相關係數、標準差及其他分別推論或估計母全體的平均數、相關及標準差，這些皆屬於統計推論之範疇。因此，我們可以說統計推論是以樣本推論母全體，以部分來推論全體，以已知來推論未知的一種統計方法。

一般的醫學或生物實驗研究上亦應用此種統計推論。依據統計學的原理，若樣本數愈大，它的分配形態與統計常數也就愈近似母體，但是，不管樣本數如何之大，祇要不等於全體數，即不可能與母全體完全相仿，此乃係抽樣變動所造成的差異。

也正因爲如此，所以凡是取自同一母體的各樣本，統計出來的各個常數或多或少總會有些差異，其差異之大小，除了與樣本大小有關外，與母全體的離差，也多少有些關連。

通常，樣本的各種統計常數中，算術平均數的變動都要比其他平均數少，而樣本差的標準差比起其他離差要可靠，因此，皆是以爲母全體的各種統計常數的最佳估計值。

爲了方便統計上的研究，樣本中的各項統計常數（ Statistics ）多以拉丁字母代之，母全體的各項統計常數（ Parmeters ）也就是理論上的算計常數，則使用希臘字母代之。例如以 M ， r ， S 來表示算術平均數、相關係數及標準差，而以 μ ， ρ ， σ 來表示母全體的算術平均數、相關係數及標準差等皆是。不過，有時也並不全然如此，爲了書寫的便利，各項統計的符號，並不依此劃分。

按各種抽樣的問題，有三種分配，一爲樣本分配（ Sample Distribution ），二爲母全體分配（ Population Distribution ），三爲抽樣分配（ Sampling Distribution ）。所謂樣本分配是指樣本內的各數值之次數分配，母全體分配係母全體內之各數值的次數分配，抽樣分配係由母全體抽取各種樣本的某種統計常數的次數分配。

假設一母全體中有N個個體，每一抽取出的樣本中有n個個體（ N ＞ n ），樣本數便有 $_NC_n$ 個，也就是

$$_NC_n = \frac{N(N-1)\cdots(N-n+1)}{n!}$$

$$= \frac{N!}{n!(N-n)!}$$

標準誤（ Standard error ），是指二次抽樣分配所得結果間的差異。各個可能樣本的平均數之標準差，叫做平均數的標準誤。標準差的抽樣分配之標準差，也就是各種可能樣本的標準差之標準差，稱爲標準差的標準誤。使用這種標準誤，可估計抽樣誤差的大小，來決定其正確性。

現今的醫學與生物實驗之研究所使用的統計推論中的抽樣分配有常態分配（ Normal distribution ）、二項分配、t 分配（ The t－distribution ）、F 分配（ The F－distribution ）、卡方分配（ The x^2 －distribution ）等分配。

應用統計推論有二種方法，一爲考驗統計假設（ Testing a statistical hypothesis ），另一爲估計母全體之常數（ Estimating population parameters ）。統計推論不是應用某一樣本常數去考驗母全體常數的假設，就是估計其母全體常數的可能的數值。

統計假設與考驗

因爲母全體的分配情形與常數都是未知數，因此在統計推論時，只能依樣本的資料來進行估計，要先提出一種假設，稱爲虛無假設（ Null

hypothesis)，通常是用H_0表示。虛無假設決定後，再考驗此假設能否成立，是接受或是摒棄。也就是對母全體的某一個未知常數，先提出一種假設，之後再考驗所得的樣本是不是在這種假設的母全體中隨機抽樣得來，決定此假設能否成立。

與之相對的假設，叫做對立假設或替代假設(Alternative hypothesis)，一般是以H_1來表示，這種假設是先假設一個母全體常數，之後再依樣本常數考驗它是否來自此母全體，若是，則此虛無假設便成立，若否，則摒棄此假設，而接受其對立假設。不過，這並不是絕對的，或許這項假設只是衆多可成立的假設之一，或許能有更接近事實的假設可成立，所以，不管虛無假設成立與否，都只是一種可能性罷了，並不是絕對的。

考驗一種假設能否成立，要以機率表示，也就是考驗一種假設，必須推算得到的樣本是得自假設的母全體的隨機抽樣的機率爲何，若可能性大，即表示樣本的常數和假設的母全體常數之間的差異，是因取樣的狀況而來，也就是被考驗的假設是可靠的。若所得樣本由假設母全體的隨機抽樣的可能性小，便表示被考驗的假設是不能成立的了。

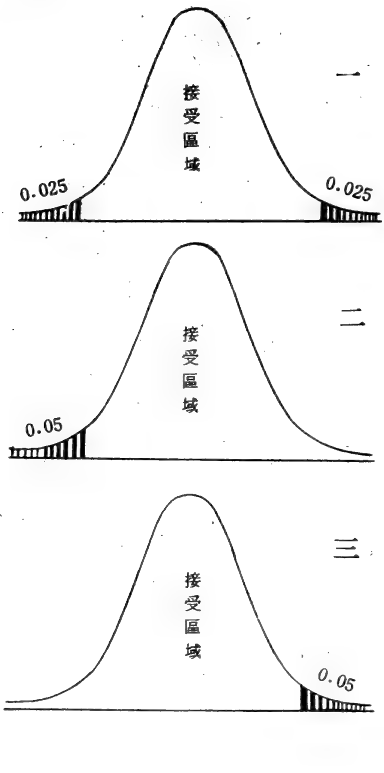
一般我們都以顯著水準(Levels of significance)來決定假設是否能成立，因此，統計假設的考驗也就稱爲顯著性考驗(Tests of significance)。考驗事實與假設之間的差異是否顯著的問題，也就是決定樣本常數及假設常數間的差異是否因機遇而影響的問題。

統計假設通常是以5％爲普通水準，1％爲高水準，0.1％爲特高水準。若一觀察之差異受抽樣變動影響的程度少於百分之一，這個假設可於1％顯著水準之下被摒棄；若少於5％時，便可在5％顯著水準下被摒棄；若少於4％，在5％顯著水準下被摒棄，但在1％顯著水準下被接受。

顯著水準低，假設被摒棄的機率大，被接受的機率則小，這樣一來

，有時眞實假設（ True Hypothesis ）有時會被否定，叫做第一種過誤（ Type I error ）。若顯著水準高，如此，假設被摒棄的機率小，被接受的機率大，有時，不眞實假設（ False Hypothesis ）也會被接受，叫做第二種過誤（ Type II error ），因此，要決定一兩全的顯著水準是不容易的，最好的方法就是折衷，取適中的5％及1％顯著水準。此外，決定顯著水準亦可視此研究對其正確性的要求之高低而酌情變動之，若摒棄原來的假設其影響不大時，可使用較低水準，接受對立假設。

若確定顯著水準後，可使得第一種過誤維持不變，再自摒棄範圍的位置來減少第二種過誤的危險，此乃因於同一水準下，可有不同之摒棄範圍的位置，可以下圖說明之：



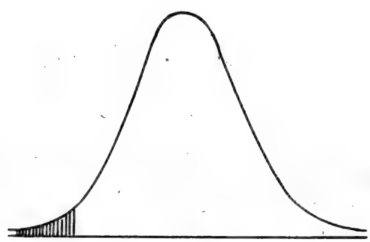
圖爲常態抽樣分配中相同顯著水準下，不同之摒棄區位置。

上圖的顯著水準是 0.05 ，此三圖的眞實假設被摒棄的過誤機會皆爲 0.05 ，但若能擇一適當之摒棄區位置，便能使過誤降至最低。可以下述三種情況，分別選一不同之摒棄區域：

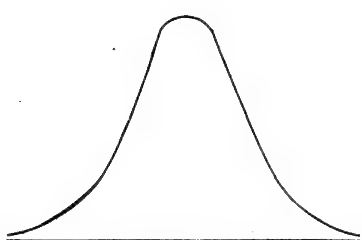
①所考驗之假設若爲 H_0 ：$\mu \geq \mu_0$ ，則對立假設爲 H_1 ：$\mu < \mu_0$ ，應取下端考驗，也就是摒棄區域集中在左邊部分。

②考驗之假設若爲 H_0 ：$\mu \leq \mu_0$ ，其對立假設則爲 H_1 ：$\mu > \mu_0$ ，應取上端考驗，也就是摒棄區域集中在右邊部分。

③考驗之假設若爲 H_0 ：$\mu = \mu_0$ ，其對立假設則爲 H_0 ：$\mu = \mu_0$ ，其對立假設則爲 H_1 ：$\mu \neq \mu_0$ ，應取二端考驗。



設 $\mu \geq \mu_0$ 時之適當摒棄區域



設 $\mu \leq \mu_0$ 時之適當摒棄區域

設$\mu=\mu_0$時之適當摒棄區域

由上述各情況，可知考驗之假設H_0：$\mu\geq\mu_0$，H_0：$\mu\leq\mu_0$或H_0：$\mu=\mu_0$，如果是在接受區域內（Accept region），則該假設便成立，否則，便應摒棄，所以，接受區域之外的部分就叫做摒棄區域（Critical region）。

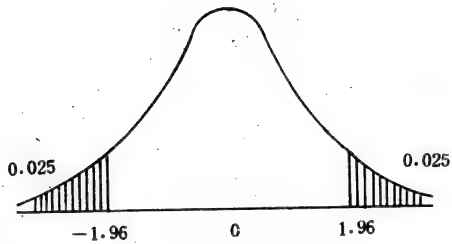
若於接受區域內的機率爲β，摒棄區域則爲$1-\beta$，也叫做a（顯著水準）。

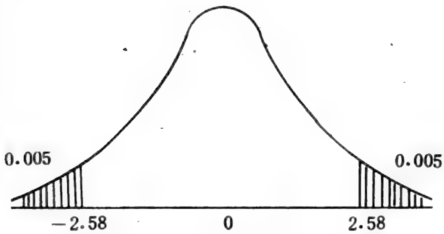
樣本數愈大，摒棄眞實假設與接受不眞實假設的可能性愈小。顯著性考驗的基本公式如下：

$$\frac{\text{樣本之統計常數}-\text{虛無假設之母全體的常數（或其他樣本常數）}}{\text{有關的標準誤}}$$

公式〔14－47〕

以常態曲線來說，公式〔14－47〕的比率，通常都以T來表示，假如結果大於或等於5％顯著水準所應有的數值（二端考驗時的5％之常態T值＝1.96），則便是顯著。假如結果大於或等於1％顯著水準所應有的數值（二端考驗時的1％之常態T值＝2.58）便是十分顯著。





圖爲顯著水準與T值對照圖

母體常數之估計

樣本常數在統計推論上，常被用來估計母全體常數（Estimating Population parameters）。這項估計涉及二個問題，一爲依據樣本常數來計算母全體常數的最佳估計值，叫做點估計（Point estimation）。如母全體之算術平均數 μ，自然是以樣本的算術平均數M來計算最爲方便，此卽爲樣本之算術平均數爲母全體之算術平均數的最佳估計值（The best single estimate）。

以樣本常數爲母全體常數的點估計，它與眞值之多少會有差異存在，其誤差之相對程度僅可使用機率方式來表示，如此，便形成了第二個範圍估計（Interval estimate）的問題，一般也叫做區間推定，也就是按機率原理，由樣本常數來決定母全體常數之所在區域範圍。

如應用樣本的算術平均數M來估計母全體的算術平均數 μ 時，其步驟爲：首先運用點估計，用已知的M爲未知的 μ 之估計值，其次再使用

區間推定，按M抽樣分配的形態和離差來確定誤差e的數值，而使原來點估計值變爲M±e之區間推定之形式。第三步驟是確定母全體平均數μ位於M－e到M＋e界限內的機率，M－e到M＋e之間的這段距離就是信賴範圍（Confidence interval），信賴範圍二端的界限M－e與M＋e叫做信賴界限。

若母全體之分配爲常態分配，樣本平均數的分配則也會是常態分配，樣本平均數分配的平均數就是母全體之平均數μ。樣本平均數的標準誤差等於樣本的次數分配之標準差除以樣本次數平方根之商（標準誤之數），依常態分配的性質，母全體平均數μ在樣本平均數和上下三個樣本平均數分配的標準差間的可能性爲99.7％。如此便可依機率原理，決定母全體平均數在某種信賴範圍下，所在的區域。

區域估計的信賴範圍，一般有99％、95％、90％、50％各種，使用最多的是95％與99％二種。樣本的大小相同、機率大小亦相同的情況下，信賴範圍愈大，估計的正確性就越小。若樣本多，可減少抽樣分配的離差，信賴範圍也就縮小，估計的值正確性便大。

樣本常數本身的性質，也足以影響統計推論的可靠性，一樣本的某一常數在抽樣分配中，若離差小，則此樣本常數對母全體常數的估計其正確程度高。由於算術平均數的抽樣分配之離差較其他平均數的抽樣分配的離差小，因此，算術平均數可做爲母全體平均數之最佳估計值。又因標準差的抽樣分配離差亦比其他離差小，因此，標準差也就是母全體差異量數的最佳估計值。

總而言之，統計推論是應用樣本常數來考驗所假設的母全體的常數（Parameter），或用來估計母全體的數值。

進行統計推論時，須依樣本常數的抽樣分配做仔細的估計。

五、卡方分配的應用

x^2 分配

x^2 分配用於樣本變異數問題與計數資料問題，以及決定樣本次數和理論次數、期望數次的功用頗廣，此分配又叫做卡方分配。

每一個自由度，都有它特殊的卡方分配，因爲每決定一個自由度 n'，便可決定一個 x^2 分配曲線。卡方分配之中心位置與分散度，隨著自由度的不同而有所變異。

下表是各不同自由度（自由度 30 以下）求出相當各機率的 x^2 值。

x^2 分配機率表

自由度	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.00004	0.00016	0.00098	0.00393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	14.9
5	0.417	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	10.6	12	14.4	16.5	18.5
7	0.989	1.24	1.69	2.17	2.83	12.0	14	16.0	18.3	20.3
8	1.34	1.69	2.18	2.73	3.49	13.4	15	17.5	20.0	22.0
9	1.73	2.05	2.70	3.33	4.17	14.7	16	19.0	21.6	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.5	21	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.70	21.1	23	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25	27.5	30.6	32.8

（續上表）

16	5.14	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	29.6	32.7	35.5	39.9	41.4
22	8.64	9.54	10.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	11.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	12.1	14.6	16.5	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	13.6	16.2	18.1	36.7	41.0	43.2	47.0	49.6
28	12.5	13.6	14.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	15.0	17.7	19.8	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
60	35.5		40.5				79.1	83.3	88.4	91.9
80	51.2		57.2				101.9	106.6	112.3	116.3
100	67.3		74.2				124.3	129.6	135.8	140.2

自由度30以上者，可依$\frac{x}{\sigma}=\sqrt{2x^2}-\sqrt{2n'-1}$公式查對常態分配機率表計算求出。

若自由度n′大於30，可先將x^2化成$w=\sqrt{2x^2}-\sqrt{2n'-1}$，然後再查對表常態機率表。

①樣本不大時母體比率P_0的x^2檢定

若自由度爲1，抽樣次數不大，可使用卡方考驗推論母全體比率P_0。應用公式如下：

$$x^2=\frac{(f-f_0)^2}{f_0}$$

公式〔14－48〕

或

$$x^2=\frac{(x-NP_0)^2}{NP_0}+\frac{(N-x-Nq_0)^2}{Nq_0}$$

公式〔14－49〕

②依樣本次數決定理論次數x^2考驗及常態性適合度的x^2考驗

$$x^2=\Sigma\frac{(f-f_0)^2}{f_0}$$

公式〔14－50〕

f：樣本次數

f_0：期望次數或理論次數

由上公式可知x^2數值的分子分母同爲正數，所以x^2值定爲正數而不爲負數。並且，x^2值亦跟著樣本的觀察次數和期望次數間的差異而改變，二者間的差異大，x^2也就大。反之則小，二者間若完全相同，x^2便爲0。自一樣本中所計算出的x^2值，該機率若在0.05和0.95之間，觀察次數和理論次數的差異可說是由機遇的影響所造成的，若機遇大於或等於95％，應審查抽樣或計算有無錯誤；若機遇小於0.05，在5％顯著水準上，放棄原假設。x^2的所在範圍，由0到∞，當$x^2=0$，即說明觀察次數及期望次數皆一致。

③二品質獨立性的 x^2 檢定

考驗二品質獨立性的 x^2，須先列出相聯表如下：

		A 品質		
		A	非 A	合計
A′品質	A′	a_{11}	a_{12}	$R_1=(a_{11}+a_{12})$
	非 A′	a_{21}	a_{22}	$R_2=(a_{21}+a_{22})$
	合計	C_1 $C_1=a_{11}+a_{21}$	C_2 $C_2=a_{12}+a_{22}$	N（即 R_1+R_2 或 C_1+C_2）

若 A 及 A′ 二品質為獨立無關聯時，則：

$$a_{11}\text{ 的理論次數}=\frac{(a_{11}+a_{12})(a_{11}+a_{21})}{N}=\frac{R_1C_1}{N}$$

$$a_{12}\text{ 的理論次數}=\frac{R_1C_2}{N}$$

$$a_{21}\text{ 的理論次數}=\frac{R_2C_1}{N}$$

$$a_{22}\text{ 的理論次數}=\frac{R_2C_2}{N}$$

相聯表原本有四格，應有四個自由度。但因為理論次數和實際次數橫的與縱的皆相等，受四個條件限制，所以失去四個自由度，但又真正失去自由度只有三個，所以只剩一個自由度。也就是原有相關表若是R列與C行，自由度便等於（R－1）（C－1）。這是指原相聯表應用絕對次數表示。若為比率表示，而對二品質的比例差異為0的顯性考驗時，只要把這個比例數變為絕對次數之後再來計算便可求出。

另有一公式亦可為此計算：

$$x^2=\frac{N((a_{11}a_{22}-a_{12}a_{21})^2)}{R_1R_2C_1C_2}$$

公式〔14－51〕

④樣本變異數推論的 x^2 檢定

$$x^2 = \frac{NS^2}{\sigma^2}$$

公式〔 14 − 52 〕

N ：樣本項數

S^2：樣本變異數

σ^2：假設母體的變異數

自由度 $n' = N - 1$

⑤成對資料差異顯著性的符號 x^2 檢定

$$x^2(1) = \frac{(n' - np)^2}{np} + \frac{(n'' - np)^2}{np}$$

公式〔 14 − 53 〕

n ：原來之對數

n'：正號之次數

n''：負號之次數

p ：$\frac{1}{2}$

⑥二資料的獨立性及相關性的聯合檢定

$$x^2 = \Sigma \frac{\left(N_{RC} - \frac{N_R N_C}{N}\right)^2}{\frac{N_R N_C}{N}}$$

$$C = \sqrt{\frac{x^2}{N + x^2}}$$

公式〔 14 − 54 〕

N ：總次數

N_R ：橫行次數的和

N_C ：縱行次數的和

N_{RC} ：相關表內方格中的次數

C：列聯係數

因爲

$$x^2=\Sigma\frac{\left(N_{RC}-\frac{N_RN_C}{N}\right)^2}{\frac{N_RN_C}{N}}$$

$$=\Sigma\frac{N_{RC}^2-2N_{RC}\frac{N_RN_C}{N}+\left(\frac{N_RN_C}{N}\right)^2}{\frac{N_RN_C}{N}}$$

$$=\Sigma\frac{N\left[N_{RC}^2-2N_{RC}\frac{N_RN_C}{N}+\left(\frac{N_RN_C}{N}\right)^2\right]}{N_RN_C}$$

$$=N\Sigma\frac{(N_{RC})^2-2N_{RC}\frac{N_RN_C}{N}+\left(\frac{N_RN_C}{N}\right)^2}{N_RN_C}$$

$$=N\Sigma\left(\frac{N_{RC}^2}{N_RN_C}-\frac{2}{N}N_{RC}+\frac{N_RN_C}{N^2}\right)$$

$$=N\Sigma\frac{N_{RC}^2}{N_RN_C}-2\Sigma N_{RC}+\frac{\Sigma N_RN_C}{N}$$

而 $\Sigma N_{RC}=N$

$\Sigma N_RN_C=\Sigma N_R\Sigma N_C=N^2$

故　$$x^2=N\Sigma\frac{N_{RC}^2}{N_RN_C}-2N+\frac{N^2}{N}$$

$$x^2=N\Sigma\frac{N_{RC}^2}{N_RN_C}-N$$

⑦中值檢定的 x^2 考驗

$$x^2 = \frac{\Sigma \frac{(n'_i)^2}{ni} - \frac{(\Sigma n'_i)^2}{\Sigma ni}}{P(1-P)}$$

公式〔14－54〕

公式〔14－54〕適合二組樣本（不成對）或k組樣本的中值檢定。方法是將各組樣本全部變量自小到大混合編排，n_i 爲 n_1 、n_2 …n_k 的各組樣本變量之個數，n_i' 爲 n_1' 、n_2' 、…n_k' ，也就是各組樣本變量中大於混合中位數的個數。

$$P = \frac{\Sigma n_i'}{\Sigma n_i}$$

六、T分配係一種單峯對稱分配

平均數抽樣分配之標準差等於 $\frac{\sigma}{\sqrt{N}}$，所以 $\frac{M-\mu}{\sigma_M}$ 爲常態分配並以0爲中心，但由於母全體的眞正標準差 σ 是未知的，所以只好用樣本的標準差S來代替，如此，計算求出之標準差只是眞正標準差的近似值，二者之間會有少數差異，若樣本數較多時，差異會較小，因此，$\frac{M-\mu}{\sigma_M}$ 還是可接近常態分配。但若是樣本數較少時，則誤差會很大，且 $\frac{M-\mu}{\sigma_M}$ 不爲常態分配，而爲t分配。

t分配（The t－Distribution）是一種單峯對稱分配，該分配的曲線型態，隨著自由度n′而變化，每一n′值便有一t分配。若n′甚小，t分配呈高狹峯，長尾巴，比常態分配的散布範圍大，它的自由度愈小，該分布愈較常態範圍大。自由度（樣本項數）增加時，便接接近常態分配。

一般情形下，當自由度n′增加到30以上時，t分配也就近似常態分配。t分配的機率值是由n′值來決定，此乃因爲t分配的曲線型態隨著自由度n′的變化而變化，所以某線上的某一段距離在曲線下所包括的面積，也隨n′的變化而不同。

下表是各不同自由度n′求得t值等於或大於某數量的機率，分列爲t分配的機率表。

t的所在範圍，理論上是從−∞到∞。

t 分配機率表

自由度	0.45	0.40	0.35	0.25	0.15	0.10	0.05	0.025	0.01	0.005
1	0.16	0.33	0.51	1.00	1.96	3.08	6.31	12.71	31.82	63.66
2	0.14	0.29	0.44	0.82	1.39	1.89	2.92	4.30	6.96	9.93
3	0.137	0.28	0.42	0.76	1.25	1.64	2.35	3.18	4.54	5.84
4	0.134	0.27	0.41	0.74	1.19	1.53	2.13	2.78	3.75	4.60
5	0.132	0.267	0.408	0.73	1.16	1.48	2.02	2.57	3.36	4.03
6	0.131	0.265	0.404	0.72	1.13	1.44	1.94	2.45	3.14	3.71
7	0.130	0.263	0.402	0.71	1.12	1.42	1.90	2.37	3.00	3.50
8	0.130	0.262	0.400	0.706	1.11	1.40	1.86	2.31	2.90	3.36
9	0.129	0.261	0.398	0.703	1.10	1.38	1.83	2.26	2.82	3.25
10	0.129	0.260	0.397	0.700	1.09	1.37	1.81	2.23	2.76	3.17
11	0.129	0.260	0.396	0.697	1.088	1.36	1.80	2.20	2.72	3.11
12	0.128	0.259	0.395	0.695	1.083	1.36	1.78	2.18	2.68	3.06
13	0.128	0.259	0.394	0.694	1.079	1.35	1.77	2.16	2.65	3.01
14	0.128	0.258	0.393	0.692	1.076	1.35	1.76	2.14	2.62	2.98
15	0.128	0.258	0.393	0.691	1.074	1.34	1.75	2.13	2.60	2.95

（續上表）

16	0.128	0.258	0.392	0.690	1.071	1.34	1.75	2.12	2.58	2.92
17	0.128	0.257	0.392	0.689	1.069	1.33	1.74	2.11	2.57	2.90
18	0.127	0.257	0.392	0.688	1.067	1.33	1.74	2.10	2.55	2.88
19	0.127	0.257	0.391	0.688	1.066	1.33	1.73	2.09	2.54	2.86
20	0.127	0.257	0.391	0.687	1.064	1.33	1.73	2.09	2.53	2.85
21	0.127	0.257	0.391	0.686	1.063	1.32	1.72	2.08	2.52	2.83
22	0.127	0.256	0.390	0.686	1.061	1.321	1.717	2.074	2.51	2.82
23	0.127	0.256	0.390	0.685	1.060	1.320	1.714	2.070	2.50	2.81
24	0.127	0.256	0.390	0.685	1.059	1.318	1.710	2.064	2.49	2.80
25	0.127	0.256	0.390	0.684	1.058	1.316	1.708	2.060	2.49	2.79
26	0.127	0.256	0.390	0.684	1.058	1.315	1.706	2.056	2.48	2.78
27	0.127	0.256	0.389	0.684	1.057	1.314	1.703	2.050	2.47	2.77
28	0.127	0.256	0.389	0.683	1.056	1.313	1.701	2.048	2.47	2.76
29	0.127	0.256	0.389	0.683	1.055	1.311	1.700	2.045	2.46	2.76
30	0.127	0.256	0.389	0.683	1.055	1.310	1.697	2.042	2.46	2.75
60	0.126	0.254	0.387	0.679	1.046	1.30	1.691	2.00	2.39	2.66
120	0.126	0.254	0.386	0.677	1.041	1.29	1.66	1.98	2.36	2.62
∞	0.125	0.253	0.385	0.674	1.036	1.28	1.65	1.96	2.33	2.58

常態分配中，似乎90 %的面積在± 1.64 標準值之間，於t 分配中；當 $n' = 10$ 時，90 % 之面積在± 1.81 標準值之間；當 $n' = 20$ 時，在± 1.72 標準值之間；當 $n' = 30$ 時，在± 1.70 之間；當 $n' = \infty$ 時，在± 1.64 之間，和常態分配完全一樣。

自由度是統計各變量中可自由變動的個數，若有一個條件來加以限制時，便會失去一個自由度。於統計推論中，作爲母全體常數之樣本自由度，即有關此估計值的獨立數值之個數。

就爲母全體標準差 σ 的估計值之樣本差 s 而言，該自由度爲（N－1），此乃因母全體各數值及其平均數之差數總和爲0，所以於應用樣本標準差 s 以估計母全體標準差 σ 時，須限制樣本差異分數的總和爲0，也就是各變量及平均數的差數和爲0），即 $\Sigma d = 0$，以使樣本平均數和母全體平均數的性質一樣，如此一來，此樣本便更能代表母全體了。對於 $\Sigma d = 0$ 的限制，要使樣本的N個差異分數中僅有N－1 個可自由獨立變化，剩下的一個差異分數受總和等於0 之限制而亦確定，便要有一個差異分數要失去其獨立變化。

換個方式來解釋此理論，由於當以樣本標準差 s 去估計母全體標準差 σ，在計算樣本標準差時，因爲母全體的平均數是未知數，不能應用母全體的差異分數 $x - \mu$，僅可利用樣本平均數的差異分數 $x - M$，如此，便須於樣本平均數及母全體平均數完全相同的條件限制下來估計母全體的標準差，這個限制便要失去一個自由度。如此便可得知，自統計資料中，每當須要確定一個常數時，就要失去一個自由度，例如計算標準差 s 時，由於要先確定一個常數（平均數），所以其自由度是N－1。再者，母全體常數的估計值的自由度，並不一定都是N－1，其自由度之減少，須看限制條件數目或所須決定的常數種類數目來決定。

自由度（Degree of freedom），原爲數學上的一個名詞，後來爲統計學應用之。舉例說明自由度的限制，如 $X + Y = 30$，若 $X = 10$

，那麼Y便爲20，兩個變量就只有一個自由度了。一變量的數值若已定，則另一變量也就因而決定了。

①平均數之差的t考驗

若兩全體變異數皆爲已知，並且兩數相等，也就是$\sigma_1^2 = \sigma_2^2 = \sigma^2$，則可應用t考驗，其公式如下：

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

公式〔14－56〕

合併標準差：

$$S_0' = \sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 - 2}}$$

公式〔14－57〕

因母全體的標準差σ未知，因此，得藉以樣本的資料來估計。公式〔14－57〕中，S_0'是σ之不偏估計值，此公式可顯示出此乃以樣本變異數S_1^2與S_2^2的加權平均數以爲σ^2之推定值，並且爲了要使其推論值爲不偏推定值，因此，用兩個樣本變異數的自由度之和（$N_1 + N_2 - 2$）作爲分母以除之。

由於平均數相差的標準誤爲：

$$\sigma_{M_1 - M_2} = S_0'\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

公式〔14－58〕

合併公式〔14－57〕與〔14－58〕得：

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 - 2}}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

公式〔14－59〕

利用 t 值來考驗母全體平均數之差是否顯著，應注意的是二個母全體必須有相當近似之變異性，方可進行推論。t 值的顯著性是否足以代表母全體的平均數的差異，應先考驗樣本的差異量數不具有顯著的差別方可。若差異量數不相等，僅比較其平均數，結果會較缺乏正確性，不足以代表母全體之差量。

②平均數的 t 值推論

$$t = \frac{M - \mu}{\dfrac{S}{\sqrt{N-1}}}$$

$$= \frac{(M-\mu)\sqrt{N-1}}{S}$$

公式〔 14 － 60 〕

均數之標準誤：

$$\sigma_M = \frac{S}{\sqrt{N-1}}$$

利用樣本平均數來推論母全體平均數，可使用 t 值機率表，先算出樣本平均數的標準值 t ，其次於應有之自由度下，便能查出機率值。（自由度爲 N － 1 ）。

③相關係數差數的顯著性之 t 考驗

$$\sigma_{z}' = \frac{1}{\sqrt{N-3}}$$

自由度＝N － 3

若樣本數增大時，將 r 化爲 Z′ 後，則極快速接近常態。

$$\sigma_{z'_1} - Z_2' = \sqrt{\frac{1}{N-3} + \frac{1}{N_2-3}}$$

$$t_{Z_1'-Z_2'}=\frac{Z_1'-Z_2'}{\sigma_{Z_1'-Z_2'}}$$

公式〔 14 － 61 〕

自由度＝N_1+N_2-6

④相關係數及廻歸係數之顯著性t考驗

$$t=\frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

公式〔 14 － 62 〕

自由度＝N－2

公式〔 14 － 62 〕僅可應用於母全體相關係數($\bar{r}$)爲0，亦即 H_0：$\bar{r}=0$ 之考驗，其他對於母全體積差相關係數假設之考驗，便須將之化爲Z′之後再予以考驗。

廻歸係數顯著性之t考驗

$$t=\frac{(b_{yx}-\bar{b}_{yx})S_x\sqrt{N-2}}{\sqrt{S_y^2-b_{yx}^2S_x^2}}$$

公式〔 14 － 63 〕

b_{yx}：樣本中Y對X之廻歸係數

$\bar{b}_{yx}$：假設的母全體Y對X之廻歸係數

自由度＝N－2，(由於$\Sigma X'=NM_x$，又$\Sigma Y'=NM_y$，因此失去二個自由度)。

下表爲統計學家費宣等人(R. A. Fisher etc)將r的5％點及1％點的t值求出，計算出相關係數後，再查表即可得知是否顯著，可省去了計算t值的手續。

相關係數之5％點及1％點表

自由度	5％點	1％點	自由度	5％點	1％點
1	0.997	1.000	24	0.388	0.496
2	0.950	0.990	25	0.381	0.487
3	0.878	0.959	26	0.374	0.478
4	0.811	0.917	27	0.367	0.470
5	0.754	0.874	28	0.361	0.463
6	0.707	0.834	29	0.355	0.456
7	0.666	0.798	30	0.349	0.449
8	0.632	0.765	35	0.325	0.418
9	0.602	0.735	40	0.304	0.393
10	0.576	0.708	45	0.288	0.372
11	0.553	0.684	50	0.273	0.354
12	0.532	0.661	60	0.250	0.325
13	0.514	0.641	70	0.232	0.302
14	0.497	0.623	80	0.217	0.283
15	0.482	0.606	90	0.205	0.267
16	0.468	0.590	100	0.195	0.254
17	0.456	0.575	150	0.159	0.208
18	0.444	0.561	200	0.138	0.181
19	0.433	0.549	300	0.113	0.148
20	0.423	0.537	400	0.098	0.128
21	0.413	0.526	500	0.088	0.115
22	0.404	0.515	1000	0.062	0.081
23	0.396	0.505			

⑤淨相關係數之t考驗

$$t = \frac{r}{\sqrt{1 - r^2}} \sqrt{N - m - 2}$$

公式〔14－64〕

r：淨相關係數

m：被固定之變數的種類數

自由度＝N－m－2＝N－K

K：變量之種類數

考驗淨相關係數r的顯著性，可查閱下表：

淨相關r的5％點及1％點表
（上表5％點，下表1％點）

自由度	變數的種類 2	3	4	5
1（5％點）	0.997	0.999	0.999	0.999
（1％點）	1.000	1.000	1.000	1.000
2	0.950	0.975	0.983	0.987
	0.990	0.995	0.997	0.998
3	0.878	0.930	0.950	0.961
	0.959	0.976	0.983	0.987
4	0.818	0.881	0.912	0.930
	0.917	0.949	0.962	0.970
5	0.754	0.836	0.874	0.898
	0.874	0.917	0.937	0.949
6	0.707	0.795	0.839	0.867
	0.834	0.886	0.911	0.927

（續上表）

7	0.666	0.758	0.807	0.838
	0.798	0.855	0.885	0.904
8	0.632	0.726	0.777	0.811
	0.765	0.827	0.860	0.882
9	0.602	0.697	0.750	0.786
	0.735	0.800	0.836	0.861
10	0.576	0.671	0.726	0.763
	0.708	0.776	0.814	0.840
11	0.553	0.648	0.703	0.741
	0.684	0.753	0.793	0.821
12	0.532	0.627	0.683	0.722
	0.661	0.732	0.773	0.802
13	0.514	0.608	0.664	0.703
	0.641	0.712	0.755	0.785
14	0.497	0.590	0.646	0.686
	0.623	0.694	0.737	0.768
15	0.482	0.574	0.630	0.670
	0.606	0.677	0.721	0.752
16	0.468	0.559	0.615	0.655
	0.590	0.662	0.706	0.738
17	0.456	0.545	0.601	0.641
	0.575	0.647	0.691	0.724
18	0.444	0.532	0.587	0.628
	0.561	0.633	0.678	0.710
19	0.433	0.520	0.575	0.615
	0.549	0.620	0.665	0.698

（續上表）

20	0.423	0.509	0.563	0.604
	0.537	0.608	0.652	0.685
21	0.413	0.498	0.552	0.592
	0.526	0.596	0.641	0.674
22	0.404	0.488	0.542	0.582
	0.515	0.585	0.630	0.663
23	0.396	0.479	0.532	0.572
	0.505	0.574	0.619	0.652
24	0.388	0.470	0.523	0.562
	0.496	0.565	0.609	0.642
25	0.381	0.462	0.514	0.553
	0.487	0.555	0.600	0.633
26	0.374	0.454	0.506	0.545
	0.478	0.546	0.590	0.624
27	0.367	0.446	0.498	0.536
	0.470	0.538	0.582	0.615
28	0.361	0.439	0.490	0.529
	0.463	0.530	0.573	0.606
29	0.355	0.432	0.482	0.521
	0.456	0.522	0.565	0.598
30	0.349	0.426	0.476	0.514
	0.499	0.514	0.558	0.591
35	0.325	0.397	0.445	0.482
	0.418	0.481	0.523	0.556
40	0.304	0.373	0.419	0.455
	0.393	0.454	0.494	0.526

（續上表）

45	0.288	0.352	0.397	0.432
	0.372	0.430	0.470	0.501
50	0.273	0.336	0.379	0.412
	0.354	0.410	0.449	0.479
60	0.250	0.308	0.348	0.380
	0.325	0.377	0.414	0.442
70	0.232	0.286	0.328	0.354
	0.302	0.351	0.386	0.413
80	0.217	0.269	0.304	0.332
	0.283	0.330	0.362	0.389
90	0.205	0.254	0.288	0.315
	0.267	0.312	0.343	0.368
100	0.195	0.241	0.274	0.300
	0.254	0.297	0.327	0.351
150	0.159	0.198	0.225	0.247
	0.208	0.244	0.270	0.290
200	0.138	0.172	0.196	0.215
	0.181	0.212	0.234	0.253
300	0.113	0.141	0.160	0.176
	0.148	0.174	0.192	0.208
400	0.098	0.122	0.139	0.153
	0.128	0.155	0.167	0.180
500	0.088	0.109	0.124	0.137
	0.115	0.135	0.150	0.162
1000	0.062	0.077	0.088	0.097
	0.081	0.096	0.106	0.115

上表對於複相關R亦可檢定，R的5％點與1％點和r完全一致。複相關之自由度＝N－K，N為樣本次數，K為變量種類數。

⑥等級相關係數之t考驗

等級相關係數的t考驗，僅須將公式〔14－62〕中之r改為ρ即可：

$$t=\frac{\rho\sqrt{(N-2)}}{\sqrt{1-\rho^2}}$$

公式〔14－65〕

ρ：等級相關係數

若不使用t值考驗，也可查閱下表，結果相同。

5％點及1％點之ρ限值表

N	5％點	1％點	N	5％點	1％點
5	0.90	1.00	16	0.43	0.60
6	0.83	0.94	18	0.40	0.56
7	0.71	0.89	20	0.38	0.53
8	0.64	0.83	22	0.36	0.51
9	0.60	0.78	24	0.34	0.49
10	0.56	0.75	26	0.33	0.47
12	0.51	0.71	28	0.32	0.45
14	0.46	0.65	30	0.31	0.43

⑦相關係數之合併

合併相關係數公式如下：

$$[Z']=\frac{\Sigma(N-3)Z'}{\Sigma(N-3)}$$

公式〔14－66〕

合併若干個樣本的相關係數時，必須先將 r 化為 Z′，之後再各乘上自由度（N－3），再將其積之和除以自由度之總和，便可求出合併後之〔Z′〕，最後再把〔Z′〕化為〔r〕即可。

七、互變異數之分析

①單因子之拉丁方格設計（Lation Square design）

此設計亦即於實驗設計時，進行二個集區設計之意。

$$ET = \frac{Vb' + Vb'' + (K-1)V_e}{(K+1)V_e}$$

公式〔14－67〕

②比較按二個標準分類樣本之相互作用

變異來源	平方和	自由度	變異數估計	F	5%點的F值	1%點的F值
b 因子	Q^2_b	$m-1$	$V_b = Q^2_b/(m-1)$	$F_1 = V_b/V_e$		
b′ 因子	$Q^2_{b'}$	$n-1$	$V_{b'} = Q^2_{b'}/(n-1)$	$F_2 = V_{b'}/V_e$		
交互作用	$Q^2_b \times b'$	$(m-1)(n-1)$	$V_b \times b' = Q^2_b \times b' / (m-1)(n-1)$	$F_3 = V_b \times b'/V_e$		
誤差	Q_e^2	$mn(k-1)$	$V_e = Q_e^2/mn(k-1)$			
總和	Q_T^2	$mnk-1$				

總變異 I：$Q_T^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$

自變數間變異（種子）：$Q_b^2 = \frac{\Sigma R^2}{nk} - \frac{(\Sigma x)^2}{N}$

自變數(肥料)間變異：$Q_{b'}^{2}=\frac{\Sigma C^{2}}{mk}-\frac{(\Sigma x)^{2}}{N}$

總變異Ⅱ：$Q_{bb'}^{2}=\frac{1}{k}\Sigma x'^{2}-\frac{(\Sigma x)^{2}}{N}$

交互作用：$Q_{b\times b'}^{2}=Q_{bb'}^{2}-Q_{b}^{2}-Q_{b'}^{2}$

誤差$=Q_{2}^{2}=Q_{T}^{2}-Q_{bb'}^{2}$

或$Q_{e}^{2}=Q_{T}^{2}-Q_{b}^{2}-Q_{b'}^{2}-Q_{b\times b'}^{2}$

③單因子之集區設計（Block design）

集區設計之意乃為把一相同特性之分子安排於若干個集區內，重複予以試驗，任一種試驗在每一集區內只有一個單位。其優點是可減少不明因素之影響，進而增加試驗的正確性。

變異來源	平方和	自由度	變異數估計	F
試驗因子	Q_{b}^{2}	$K-1$	$V_{b}=Q_{b}^{2}/(K-1)$	$F_{1}=V_{b}/V_{e}$
集　區	$Q_{b'}^{2}$	$K'-1$	$V_{b}'=Q_{b}'^{2}/(K'-1)$	$F_{2}=V_{b}'/V_{e}$
誤　差	Q_{e}^{2}	$(K-1)(K'-1)$	$V_{e}=Q_{e}^{2}/(K-1)(K'-1)$	
總　和	Q_{T}^{2}	$N-1$		

集區設計之功能公式如下：

$$ET=\frac{(K'-1)Vb'+K'(K-1)V_{e}}{V_{e}(KK'-1)}$$

公式〔14－68〕

④樣本數少時之母體比率假設之F考驗及比率估計

ⓐ母體比率假設之考驗

樣本比率大於母體之假設比率（P_{0}）時

$$F=\frac{n_2'(1-P_0)}{n_1' P_0}$$

$$=F(n_1', n_2')$$

公式〔14－69〕

$n_1'=2(n-k+1)$

$n_2'=2k$

n：樣本次數

k：樣本某一相同特性出現次數

若 H_0 ：$P_0 \leq P_0'$，放棄假設之範圍爲 $F > F_a$

若 H_0 ：$P_0 = P_0'$，放棄假設之範圍爲 $F > F_{\frac{a}{2}}$

樣本比率小於母體之假設比率（P_0）時

在同一行或同一列中，相同的實驗僅能有一次，若同行僅是出現一次，同列便可出現二次或者更多次，抑或相反的，同列只出現一次，而同行出現二次或以上，這便爲一種集區設計。

進行拉丁設計之目的在於求得最爲正確有效之試驗結果，除去原因未明之因素，此以下圖說明之：

A	A	A	A
B	B	B	B
C	C	C	C
D	D	D	D

上圖即爲一種集區設計，但有時一些因素會導致試驗結果正確性的低弱，若採用拉丁方格設計，便不易產生失誤。如下：

A	B	C	D		A	B	C	D		
B	D	A	C	或	B	A	D	C	等。
C	A	D	B		C	D	A	B		
D	C	B	A		D	C	B	A		

變異來源	平方和	自由度	變異數估計	F
試驗因子	Q^2b	$K-1$	$Vb=Q^2b/(K-1)$	$F_1=Vb/Ve$
行間間	Q^2b'	$K-1$	$Vb'=Q^2b'/(K-1)$	$F_2=Vb'/Ve$
列間	Q^2b''	$K-1$	$Vb''=Q^2b''/(K-1)$	$F_3=Vb''/Ve$
誤差	Q^2e	$(K-1)(K-2)$	$Ve=Qe^2/(K-1)(K-2)$	
總和	Q^2T	$N-1$		

拉丁方格設計之功能如下公式：

$$F'=\frac{n_2'' P_0}{n_1''(1-P_0)}$$

$$=F'(n_1'', n_2'')$$

公式〔14－70〕

$n_1''=2(K+1)$

$n_2''=2(n-K)$

若 H_0 ：$P_0 \geq P_0'$，放棄假設之範圍為 $F'>F'_a$

若 H_0 ：$P_0 = P_0'$，放棄假設之範圍為 $F'>F'_{\frac{a}{2}}$

ⓑ母體比率之估計

$$P_{OL}=\frac{n_2'}{n_2'+n_1'F_0}$$

公式〔14－71〕

$$P_{ov}=\frac{n_1''\,F_0'}{n_2''+n_1''\,F'_0}$$

⑤迴歸的直線性考驗

依相同事實所求出的直線相關係數及曲線相關比，經過F考驗之後，該趨向一致。若二變量之間有關聯存在，其相關是為直線或曲線者，也可以使用F值來考驗。考驗時的虛無假設，爲二變量之相關爲直線的，F值大者，是曲線相關，F值小者是直線相關。

$$F_{(K-2,N-K)}=\frac{\dfrac{\eta_{yx}^2-r^2}{K-2}}{\dfrac{1-\eta_{yx}^2}{N-K}}$$

（K爲X之組數）

公式〔14－72〕

⑥相關係數及相關比的顯著性F考驗

F考驗可用來証明天地萬物間，任二者事物或現象的關係。宇宙中，無論何種事物或現象，它們之間的積率相關或相關比，幾乎不可能爲0，或多或少總會有些相關聯，欲了解其彼此間是不是有關聯，或是關聯的程度大小，可應用F考驗來証明。

$$F_{(K-1,N-K)}=\frac{\dfrac{r^2}{K-1}}{\dfrac{1-r^2}{N-K}}$$

$$=\frac{r^2\,(N-2)}{1-r^2}$$

公式〔14－73〕

由於上公式中的K爲2，因此可變爲下式：

$$F_{(k-1, N-k)} = \frac{\dfrac{\eta^2_{yx}}{K-1}}{\dfrac{1-\eta^2_{yx}}{N-K}}$$

公式〔14－74〕

K 爲縱行數，X 之組數

⑦複相關R 的顯著性F 考驗

其公式如下：

$$F_{(k-1, N-k)} = \frac{R^2}{K-1} \cdot \frac{N-K}{1-R^2}$$

公式〔14－75〕

⑧二樣本變異數差異的顯著性之F 考驗

此考驗亦即二母群體變異數假設之考驗，可使用下列公式：

$$F(n_1', n_2') = \frac{N_1(N_2-1)S_1^2}{N_2(N_1-1)S_2^2} \quad \frac{\sigma_2^2}{\sigma_1^2}$$

公式〔14－76〕

或

$$F(n_1', n_2') = \frac{\dfrac{\Sigma X_1^2}{N_1-1}}{\dfrac{\Sigma x_2^2}{N_2-1}} \quad \frac{\sigma_2^2}{\sigma_1^2}$$

公式〔14－77〕

$n_1' = N_1 - 1$

$n_2' = N_2 - 2$

假設二母群體之變異數相同，也就是H_0：$\sigma_1^2 = \sigma_2^2$　時，$\dfrac{\sigma_2^2}{\sigma_1^2} = 1$，則公式後之$\dfrac{\sigma_2^2}{\sigma_1^2}$便可省略不予計算。

⑨按二個標準分類的樣本之F考驗

也就是二個實驗因子（自變數）和一個因應數間所產生差異的顯著性考驗。舉例言之，用三種不同的種子，加以四種的肥料，於 12 塊土質一樣、面積大小相同的土地中進行實驗，此研究中，種子和肥料皆爲自變數，產量爲因變數。在計算各種變異數時，可使用下列公式：

總變異：$Q_T^2 = \Sigma(x - \overline{x})^2$

公式〔14－78〕

種子間變異：$Q_b^2 = \Sigma k_t'(\overline{x}_R - \overline{x})$

$$= \frac{\Sigma k_R^2}{k_j'} - \frac{(\Sigma x)^2}{N}$$

公式〔14－79〕

肥料間變異：$Q_{b'}^2 = \Sigma k_1(\overline{x}_C - \overline{x})^2$

$$= \frac{\Sigma k_C^2}{k_1} - \frac{(\Sigma x)^2}{N}$$

公式〔14－80〕

誤　差：$Q_e^2 = Q_T^2 - Q_b^2 - Q_{b'}^2$

公式〔14－81〕

⑩按一個標準分類的三個以上樣本平均數之差量的顯著性考驗

若於研究中，除了因變數外，只有一種自變數（也就是只有一種實驗因子），在分析研究結果時，以自變數爲分類標準，而比較因變數的變異是不是顯著。

分析變異數的程序，可以下表來說明：

變異來源	離均差平方和	自由度	變異數估計	F
組　間	$Q_b^2=\Sigma K_i(\overline{X}_i-\overline{X})^2$	$K-1$	$S_b^2=\frac{Q_b^2}{(K-1)}$	$F(K-1)(N-K)=\frac{S_b^2}{S_w^2}$
組　內	$Q_w^2=\Sigma\Sigma(X_i-\overline{X}_i)^2$	$N-K$	$S_w^2=\frac{Q_w^2}{N-K}$	
總變異	$Q_T^2=\Sigma(X-\overline{X})^2$	$N-1$	$S_T^2=\frac{Q_T^2}{N-1}$	

$Q_b^2=\Sigma K_i(\overline{X}_i-\overline{X})^2$ 表示各組間之變異

K_i 爲 i 組之項數

$\overline{X}_i$ 爲 i 組之平均數

$\overline{X}$ 爲總平均數

此變異可能由於部分的機遇誤差，部分的自變數（實驗因子）所得。

$Q_w^2=\Sigma\Sigma(X_i-\overline{X}_i)^2$ 表示各組內之差異，此變異乃爲機遇誤差或試驗誤差，與自變數（實驗因子）無關。

$Q_T^2=\Sigma(X-\overline{X})^2$ 表示總變數，爲組間變異及組內變異之和，即 $Q_T^2=Q_b^2+Q_w^2$，總變異乃由於機遇誤差而來，也與實驗因子及其他因素有關。

估計組間之變異數，由於要先決定全體變量之總平均數 $\overline{X}$，因此，失去一個自由度，而以（$K-1$）爲分母。

估計組內變異數，由於要先決定K個平均數，因此失去K個自由度，因而以 $N-K$ 爲分母。

總變異之公式：

$$\Sigma fd'^2-\frac{(\Sigma fd')^2}{N}$$

公式〔14－82〕

$$組間變異=\Sigma\frac{(\Sigma f_i d_i')^2}{K_i}-\frac{(\Sigma fd')^2}{N}$$

公式〔14－83〕

$$組內變異=\Sigma fd'^2-\Sigma\frac{(\Sigma f_i d_2')^2}{K_i}$$

公式〔14－84〕

互變異數分析

以上所討論之變異數分析方法，包含有二種變異，其一爲由已知原因（實驗因子）所產生的變異，其二爲由未知原因及機遇誤差所產生的變異。

爲了使得研究實驗的結果能更加正確，必須將實驗中所產生的不明因素的影響排除，要排除此因素，一般有二種方法，其一是，於實驗之初詳細的檢驗及設計，小心挑選實驗之對象，實驗中，適宜的控制各種情況，使得各組之中，除了實驗的因子不同外，各種情況皆須相同。其二就是互變數分析法，此互變數分析法乃將和未知因素關係密切之數作爲自變數，實驗結果作爲因變數，再配合廻歸直線之方法，由實驗結果中，去掉自變數所代表的未知因素的影響，由於此乃依二種或二種以上的變量數列來分析，因此叫做互變異分析法。

應用互變異分析法有二項限制，一爲除了實驗之結果外，並需得知其與未知因素相關之數列；二爲，以爲變異數分析之資料，可利用廻歸直線處理，依自變量將試驗結果加以修正，再按修正後之結果來進行變異分析，此二項爲互變異分析之條件。

假設X是自變數，Y是因變數，使用互變異數分析法考驗K個調整後之平均數之間的顯著性爲何，其程序如下：

①應用變異數分析法將x及y之全部平方和分解爲二個組成分子，一

爲組內平方和，另一爲組間平方和。

②把 xy 的全部乘積和分解爲二個組成分子，一爲組內乘積和，另一爲組間乘積和。

③求出調整過的 y 之總平方和，用來去除互變異數 x 之直線影響。

④運用組內之 y 對 x 的廻歸，求出修正過之 y 的組內平方和。

⑤由調整過之總平方和中減去調整過之組內平方和，可求出調整過之組間平方和。

⑥以自由度 $n_1' = K - 1$ 除調整過之組間平方和，可求出變異數估計值 S_b^2，以自由度 $n_2' = N - K - 1$ 除調整過之組內平方和，可求得變異數估計值 S_w^2。

⑦求出 F 值 $= \frac{S_b^2}{S_w^2}$。

⑧查表，考驗調整過之 Y 的平均數間的顯著性。

一群體之中，每個個體之間總會有一差異，此差異爲變異。這是由於某一已知或未知的因素而造成的，也可能是由於抽樣之機遇而產生。變異數也就是標準差之平方，總變異中含有二種變異，一是已知原因的變異，另一是試驗變異。

所謂變異數分析，即是將樣本之總變異數分析作爲已知原因所造成的變異數及實驗變異數。在科學實驗研究中，考驗某種原因所引起之變異是否顯著是十分重要的。若實驗因素只有二種，可先比較二平均數的差異後依常態分配或 t 分配作顯著性考驗。母全體的變異數之間是否不具有顯著差異，小樣本時須以 F 分配來作變異數分析，方可明瞭。若是驗因素爲三個或者更多時，各個平均數的差異是否顯著，則須利用變異數分析方可明瞭。除此之外，考驗相關係數、相關比、複相關係數及廻歸方程的顯著性或直線性，也都須利用變異數分析。

假設 σ^2 爲一常態母全體之變異數，首先取一隨機樣本，此樣本爲 N_1，則母全體變異數 σ^2 的不偏估計值 S_1^2 便爲：

$$\frac{\Sigma(X_1 - M_1)^2}{N_1 - 1}$$

或

$$\frac{\Sigma x_1^2}{N_1 - 1}$$

另取一隨機樣本，此樣本為N_2，此時母全體變異數σ^2的另一個不偏估計值S_2^2便為為：

$$\frac{\Sigma x_2^2}{N_2 - 1}$$

這二個變數的比例就叫做 F 比例。

$$F = \frac{\dfrac{\Sigma x_1^2}{(N_1 - 1)}}{\dfrac{\Sigma x_2^2}{(N_2 - 1)}}$$

$$= \frac{S_1^2}{S_2^2}$$

公式〔 14 － 85 〕

由公式〔 14 － 85 〕中可知一母全體變異數的二個不偏估計值之間的比例即為 F 比例。F 比例抽樣分配之形態乃隨著分子分母自由度n_1'及n_2'在變化，各對之自由度可獲一 F 統計量，此統計量可以$F(n_1', n_2')$來表示。由於分子及分母之自由度乃依限制的條件來決定，因此公式〔 14 － 84 〕可變化如下：

$$F(n_1', n_2') = \frac{\dfrac{\Sigma x_1^2}{n_1'}}{\dfrac{\Sigma x_2}{n_2'}}$$

公式〔 14 － 86 〕

F之值之所以爲正值，此乃因分子分母皆爲正值之故，因此不會有負值出現。又由於F之範圍爲0到∞之間，因此 n'_1 及 n_2' 若爲無限大時，F分配也不能用常態爲極限。若化爲 $Z = \log_e F$ 時，倘 n'_1 ，n_2' 爲無限大，此時，Z之分配卽可用常態分配作爲極限。

須應用各不同重要機率之F值，可查閱下表：

F分配機率表

小均方自由度	機率 \ 大均方自由度 F值	1	2	3	4	5	6	7	8	9	10	20	30	60	120	∞
1	0.05	161	200	216	225	230	234	237	239	241	242	248	250	252	253	254
	0.025	648	800	864	900	922	937	948	957	963	969	993	1000	1010	1015	1020
	0.01	4052	5000	5404	5630	5764	5860	5930	5980	6022	6060	6210	6260	6310	6340	6370
	0.005	16200	20000	21600	22500	23100	23400	23700	23900	24100	24200	24800	25000	25300	25400	25500
2	0.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5
	0.025	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
	0.01	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
	0.005	198	199	199	199	199	199	199	199	199	199	199	199	199	199	200
3	0.05	10.1	95.5	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.62	8.57	8.55	8.53
	0.025	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.2	14.1	14.0	14.0	13.9
	0.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	26.7	26.5	26.3	26.2	26.1
	0.005	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	42.8	42.5	42.1	42.0	41.8
4	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.75	5.69	5.66	5.63
	0.025	12.3	10.6	9.93	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.56	8.46	8.36	8.31	8.26
	0.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.0	13.8	13.7	13.6	13.5
	0.005	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.2	19.9	19.6	19.5	19.3
5	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.50	4.43	4.40	4.36
	0.025	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.33	6.23	6.12	6.07	6.02
	0.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.55	9.38	9.20	9.11	9.02
	0.005	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	12.9	12.7	12.4	12.3	12.1
6	0.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.81	3.74	3.70	3.67
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.17	5.07	4.96	4.90	4.85
	0.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.40	7.23	7.06	6.97	6.88
	0.005	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	9.59	9.36	9.12	9.00	8.88
7	0.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.38	3.30	3.27	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.47	4.36	4.25	4.20	4.14
	0.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.16	5.99	5.82	5.74	5.65
	0.005	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	7.75	7.53	7.31	7.19	7.08

8	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.08	3.01	2.97	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.00	3.89	3.78	3.73	3.67
	0.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.36	5.20	5.03	4.95	4.86
	0.005	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	6.61	6.40	6.18	6.06	5.95
9	0.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.86	2.79	2.75	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.67	3.56	3.45	3.39	3.33
	0.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	4.26	4.81	4.65	4.48	4.40	4.31
	0.005	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.53	6.42	5.83	5.62	5.41	5.30	5.19
10	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.70	2.62	2.58	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.42	3.31	3.20	3.14	3.08
	0.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.41	4.25	4.08	4.00	3.91
	0.005	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.27	5.07	4.86	4.75	4.46
20	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	2.04	1.95	1.90	1.84
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.46	2.35	2.22	2.16	2.09
	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	2.94	2.78	2.61	2.52	2.42
	0.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.32	3.12	2.92	2.81	2.69
30	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.84	1.76	1.68	1.62
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.20	2.07	1.94	1.87	1.79
	0.01	7.58	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.55	2.39	2.21	2.11	2.01
	0.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	2.82	2.63	2.42	2.30	2.18
40	0.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	3.08	1.84	1.74	1.64	1.58	1.51
	0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.07	1.94	1.80	1.72	1.64
	0.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.37	2.20	2.02	1.92	1.80
	0.005	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.60	2.40	2.18	2.06	1.93
60	0.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.75	1.65	1.53	1.47	1.39
	0.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	1.94	1.82	1.67	1.58	1.48
	0.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.20	2.03	1.84	1.73	1.60
	0.005	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.39	2.19	1.96	1.83	1.69
120	0.05	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.66	1.55	1.43	1.35	1.25
	0.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.82	1.69	1.53	1.43	1.31
	0.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.03	1.86	1.66	1.53	1.38
	0.005	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.19	1.98	1.75	1.61	1.43
∞	0.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.57	1.46	1.32	1.22	1.00
	0.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.71	1.57	1.39	1.21	1.00
	0.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	1.88	1.70	1.47	1.32	1.00
	0.005	7.88	5.30	4.38	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.00	1.79	1.53	1.36	1.00

上表僅列舉F值之右邊尾部機率，左邊之機率可以倒數計算出，例如，若要查F（10.6）的右邊面積百分比時，可於上表中橫行10，縱行6之位置查出F＝4.06的右邊面積爲5％。

全部之F值，一般叫做危險點（Critical points）。若爲一端考驗，F（10.7）分配中上端5％摒棄區域之危險點爲3.64，下端5％摒棄區域的危險點卽爲$\frac{1}{3.14}$，也就是0.32。若爲二端考驗，F(10.7)分配中，二端5％摒棄區域之危險點爲4.76及$\frac{1}{3.95}$。由此可知，若要求出F（n'_1，n_2'）分配中的左端危險點，可自F（n_2'，n_1'）分配的右端危險點的倒數求得。若欲求之數值於表中查不到時，可以插補法或近似值來代替。

第十五章
計數資料(卡方檢驗)的應用

假使有一母全體，須由樣本統計來估計，此爲對於方格次數（Cell frequencies）的一種限制，因此，必須在獨立之方格次數（independent cell frequencies）的數目裏減去1。一般在求適合度檢定（goodness - of fit test）時應用的卡方考驗之自由度時，由於各方格的次數之和須與樣本的總次數相等，必須要減掉一個自由度。

除此之外，由於群體常數須由樣本統計數來估計，因此又必須將各個未知的群體常數各減去一個自由度。

若自由度爲1時，分配之近似法要比其他自由度不精準，要消除此缺憾，可將各f－F之值先減掉0.5再予以平方，可得較正確之值。因此可改公式〔15－1〕爲公式〔15－2〕：

$$x^2 = \frac{\Sigma(f-F)^2}{F}$$

公式〔15－1〕

$$x^2 = \Sigma\frac{(|f-F|-0.5)^2}{F}$$

公式〔15－2〕

非獨立樣本比例數之卡方考驗（chi - square test for nonindependent sample proportions），又可稱爲McNemar氏考驗（McNemar's test）。公式如下：

$$x^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

公式〔 15 − 3 〕

若應用連續性校正，則可改公式爲：

$$x_c^2 = \frac{(|n_{12} - n_{21}| - 1)^2}{n_{12} + n_{21}}$$

公式〔 15 − 4 〕

卡方考驗不但可適用於單一的隨機樣本，同時也適用於二個以上獨立之隨機樣本（ independent random sample ）。此考驗可使研究者知其是否有對資料作更深入分析的價值。此外，應用卡方考驗須注意二點：其一爲資料須爲點計形式（ in the form of cocents ），其二爲研究之對象之間，於統計學上，須爲獨立性（ The observations must be ttatistically independent ）。

第十六章
計數資料(比例數)的應用

計數資料（enumeration data）乃指討論來自群體的樣本時，以點計歸屬於一類的數目即稱。如此之資料，只有二種可能，一爲「正面」，一爲「負面」。我們只要記錄樣本中的任何一被研究對象之屬性爲何即可。

通常以 π 來作爲群體比例數（population proportion）的符號，資料中之群體常數（population parameter）是被研究對象的特定屬性（particular attribute）於群體中所占之比。樣本比例數以P來代表，是樣本中所能得到的對 π 之最佳估計值（best estimate）。

此章所探討的P之分配情形是求出 π 的信賴範圍及考驗統計假設。

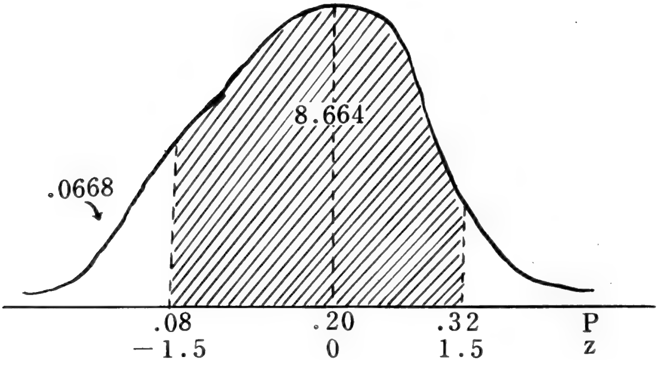
若樣本數極大，則樣本比例數的分配也就越接近常態分配。當樣本足夠大時，常態表可用爲二項分配的近似法。依照先前之研究慣例，每當 $n\pi$ 及 $n(1-\pi) \geq 5$，即可以常態曲線來求出本樣本比例數的問題。此慣例告訴我們，當 $\pi = 0.5$，比 π 近於0或1時更似常態曲線。

使用常態分配近似法（normal distribution approximation）以替代樣本比例數之分配（distribution of sample proportions），以下例說明之。

在服用一藥物的人的群體中，抽取大小爲 25 的樣本，設該群體患者中有一特定副作用的所占比例數爲 .2，那麼，在樣本大小爲 25 的所有可能之樣本比例數中，介於 .08 到 .32 之間的比例數有若干？

下圖可作爲解答此問題的參考。

P 介於 .08 ～ .32 中者所占之比例



由於 $n\pi = 25(.2) = 5$，$n(1-\pi) = 25(.8) = 20$ 均大於或等於5，因此可應用常態分配近似法。全部樣本比例數的均數 μ_p 爲.2，上圖中之中打斜線之區域卽爲一近似之答案。於應用常態曲線近似法時，所使用的常態曲線、均數爲 .2，標準差 σ_p 爲：

$$\sigma_p = \sqrt{\pi(1-\pi)/n}$$

$$= \sqrt{.2(1-.2)/25}$$

$$= \sqrt{.16/25}$$

$$= 0.8$$

若要求出介於 .08 ～ .32 之間的面積，應先求出相當於 .32 的Z值，此爲：

$$Z = (P - \mu_p)/\sigma_p$$

$$= (.32 - .20)/.08$$

$$= .12/.08$$

$= 1.5$

由表可知，9332（93.32％）的樣本比例數皆於.32的左側，所以6.68％的樣本比例數大於.32。同理，由於6.68％的樣本比例數小於.08，因此 $1-2X\ \ .0668=0.8664$（86.64％）的樣本比例數介於.08～.32之中。

樣本數爲 25 的全部可能樣本比例數中，大約有86.64％皆介於.08～.32之中。若以二項分配解之，所求出的結果爲.9258（92.58％）。本例的 $n\pi=5$，勉強可算是達到了可取之邊緣，因此，常態近似法不是很接近。此缺憾可以連續性校正法來補救。當變換成Z時，便以 $P+\frac{1}{2n}$ 代替P，爲：

$$P+\frac{1}{2n}=.32+\frac{1}{2\times 25}$$

$$=.32+\frac{1}{50}$$

$$=.32+0.2$$

$$=.34$$

因此

$$Z=\frac{\left(P+\frac{1}{2n}\right)-\mu_p}{\sigma_p}$$

$$=\frac{\left(.32+\frac{1}{2\times 25}\right)-20}{\sqrt{0.2(1-.2)/25}}$$

$$=\frac{(.32+.02)-.20}{.08}$$

$$=\frac{.14}{.08}$$

$$=1.75$$

由表知 .9599（ 95.99％ ）的樣本比例數皆小於 .34 ，也就是說 .9599 的樣本比例數等於或小於 .32 ，由於當樣本 $n = 25$ 時，不會有樣本比例數 P 之值介於 .32 ～ .36 之中，所以 4.01％ 的樣本比例數大於 .32 及 4.01％的樣本比例數小於 .08 。介於 .08 ～ .32 之間的樣本比例數共有 91.98％。

在討論兩均數之差時，曾說過，若 $\bar{x}_1$ 爲常態分配，其均數爲 μ_1 ，變異數爲 $\sigma^2_{\bar{x}_1}$ ，若 $\bar{x}_2$ 也爲常態分配，其均數爲 μ_2 ，變異數爲 $\sigma^2_{\bar{x}_2}$，則 $\bar{x}_1 - \bar{x}_2$ 也就爲常態分配，其均數爲 $\mu_1 - \mu_2$ ，變異數 $\sigma^2_{\bar{x}_1 - \bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}$ 。

假使 n_1 及 n_2 之數皆足夠大，那麼此二比例數之差的情形便差不多和二均數的差完全一樣。

如果樣本數夠大，那麼二比例數的差數呈常態分配，它的均數等於二群體比例數的差，變異數等於 P_1 及 P_2 的二個變異數的和。也就是 P_1 爲常態分配，均數爲 π_1 ，變異數爲 $\pi_1(1-\pi_1)/n_1$ 。P_2 爲常態分配，均數爲 π_2 ，變異數爲 $\pi_2(1-\pi_2)n_2$ 。則 $P_1 - P_2$ 也爲常態分配，均數爲 $\pi_1 - \pi_2$ ，變異數（ $\sigma^2_{p_1-p_2}$ ）爲 $\pi_1(1-\pi_1)/n_1 + \pi_2(1-\pi_2)/n_2$ 。

由於 $P_1 - P_2$ 分配的標準差爲 $\sqrt{\pi_1(1-\pi_1)/n_1+\pi_2(1-\pi_2)/n_2}$ ，則計算 $\pi_1 - \pi_2$ 的 95％ 信賴範圍的方法如下：

$$(P_1-P_2)\pm 1.96\sqrt{\pi_1(1-\pi_1)/n_1+\pi_2(1-\pi_2)/n_2}$$

二樣本比例數之差之信賴範圍公式：

$$(P_1-P_2)\pm 1.96\sqrt{P_1(1-P_1)/n_1+P_2(1-P_2)/n_2}$$

如果樣本數足夠大，對於群體比例數之假設可應用常態曲線同群體均數之假設一樣來行統計考驗（ statistical tests ）。

當我們在作研究考驗時，常會發生同一組資料在二種不同的考驗方

法下，會有二種不同的研究結果，這告訴我們，在行考驗之前便要審慎擇一可靠且正確性高的方法予以考驗。

來自二個不同全群的二個樣本比例數計算出以後，有些研究者常以考驗假設來決定二比例數的差是否爲0，而不作群體比例數的信賴範圍。

若使用常態表來變換爲Z值，需先求出 $P_1 - P_2$ 分配的標準差，其公式如下：

$$\sigma_{p_1-p_2} = \sqrt{\pi_1(1-\pi_1)/n_1+\pi_2(1-\pi_2)/n_2}$$

第十七章
群體變異數估計的應用與重要性

無論我們所從事研究的對象爲何，最重要的，就是研究它的變異性，欲求算估計全體均數，亦須先行估計其變異數及標準差，之後再使用此估計值來考驗群體均數的假設及群體均數之信賴範圍。

若樣本僅有一個，則此樣本的變異數 S^2 便可作爲群體變異數 σ^2 最佳之點估計（best point estimatf），此外，樣本標準差 S 亦可作爲群體標準差 σ 之點估計。但常會有二個以上的樣本可估計其變異數，如此，我們便可將二個樣本變異數合併，以 S_p^2 來估計 σ^2，其公式如下：

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

公式〔17－1〕

又假使有若干個樣本（以 K 表示），同樣可作如下之合併：

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2+\cdots\cdots+(n_k-1)S_k^2}{n_1+n_2+\cdots\cdots+n_k-K}$$

公式〔17－2〕

K 的樣本之變異數各爲 S_1^2，S_2^2，$S_3^2 \cdots S_k^2$，樣本大小各爲 n_1，n_2，$n_3 \cdots n_k$。

由一個樣本計算群體變異數及標準差之信賴範圍。

如果資料乃來自一呈常態分配之群體中，卡方表便能被利用來計算單一群體變異數的信賴範圍。自一變異數爲 σ^2 的常態群體中隨機抽取

樣本n，若樣本之變異數爲S^2已知，便可計算出量數$(n-1)S^2/\sigma^2$，此量數乃隨樣本的不同而變化，因此有一分配，這個分配也就是卡方分配（chisquare distribution），自由度爲$n-1$，也就是計算S^2時所使用的分母。

以二個以上的樣本計算群體變異數及標準差之信賴範圍。

若估計一群體變異數σ^2的樣本有二個以上，便可使用合併變異數S_p^2來求一信賴範圍，其公式爲：

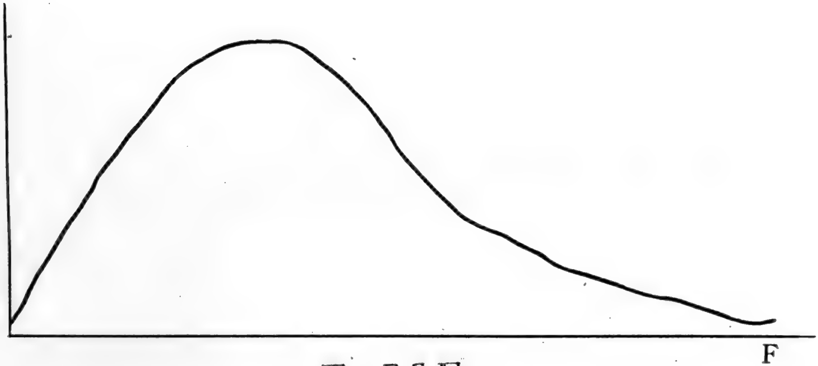
$$\frac{(n_1+n_2+\cdots n_k-K)S_p^2}{\sigma^2}$$

公式〔17－3〕

若全部之樣本皆來自於常態群體，同時變異數σ^2也一致，則自由度爲$n_1+n_2+\cdots n_k-K$，所使用的自由度也就是計算S_p^2時之分母。

若欲只用一種統計考驗來判斷二個樣本是否來自變異數相同之群體，可應用t考驗來檢視二組未配對資料的均數是不是相等。要假設二群體的變異數相等，須先作初步考驗，証明是否二群體之變異數眞相等。

若有大小分別爲n_1與n_2的二個獨立隨機樣本，便可計算其樣本變異數S_1^2及S_2^2。假使自此抽取樣本的二個群體皆爲常態分配，變異數各爲σ_1^2及σ_2^2，便可以數學方法來驗証$(S_1^2/\sigma_1^2)/(S_2^2/\sigma_2^2)$爲F分配，此數之分子若用自由度$n_1-1$來乘，便爲卡方分配，分母亦同。至於F分配的形狀乃視其分子與分母之自由度n_1-1及n_2-1來決定，大致如下圖：



圖：F 分配

σ_1^2 若比 σ_2^2 大很多，S_1^2 也會比 S_2^2 大許多，如此，F 即可能甚大。同樣的，σ_1^2 如果比 σ_2^2 小很多，F 便可能會相當之小。所以，若求出之 F 值甚大或甚小，我們便可知此二群體變異數不等，應作雙邊考驗。

第十八章
估計及信賴範圍的定義及應用

信賴範圍的決定，乃是於多次的反複抽樣中以一種獨特的方式自樣本中求取的一段範圍，如此求取之信賴範圍勢必會有95％的信賴範圍會含括群體常數的值在內。

信賴範圍是依據研究者對研究對象之準確性的多寡的要求而決定，並不一定要爲95％，信賴水準愈高，則準確性也就愈高，這是相對的，一般常用的信賴水準有90％，95％和99％三種，通常研究者都選擇95％的信賴水準。

增加樣本數可使之達到信賴範圍之區間短而信賴水準又高的要求。

樣本大小的決定亦無一定的標準，不同的研究對象所需之樣本大小互異。無所謂適用於任何研究對象之最佳樣本數。樣本的大小除須視研究對象爲何外，群體變異數之大小，期望之信賴範圍爲何，信賴水準又若干皆爲決定樣本大小應考慮之因素。

在作考驗時，變異數常爲未科數，因此必須使用樣本資料來估算，可利用$Z=(\bar{x}-\mu)/(0/\sqrt{n})$呈常態分配的事實。一般情形中，$\sigma$值也常爲未知數，因此須先求出S值以估算$\sigma$，所以可以$t=(\bar{x}-\mu)/(S/\sqrt{n})$的一數量來替代$Z=(\bar{x}-\mu)/(\sigma/\sqrt{n})$。

t分配與常態分配很近似，乃呈鐘形對稱，比標準常態分配的散佈廣。t分配乃隨樣本數大小的變化而變化，當n值小時，t分配末端所占之面積較大。當n值大時，曲線之散佈便小，而會很近似常態分配。

若由樣本之S來估計標準差，則與σ爲已知一樣，也能求群體均數

μ之信賴範圍，差別只是S代替了σ，且t表代替了常態表。

若σ爲已知，μ的95％之信賴範圍爲$\overline{x} \pm Z(.975)\sigma/\sqrt{n}$。Z(.975)爲常態曲線全面積的97.5％皆於Z值的左側。若由樣本求出的S代替σ，則μ之95％信賴範圍爲$\overline{x} \pm t(.975)S/\sqrt{n}$。t(.975)爲d.f.$=n-1$的t分配曲線全面積的97.5％皆於t值之左側。

在μ_1及μ_2的均數中與變異數爲σ^2的常態群體中抽取大小爲n_1與n_2的樣本，可以數學方法証明$\overline{x}_1-\overline{x}_2$呈常態分配，均數是爲$\mu_1-\mu_2$，變異數爲$\sigma^2(1/n_1+1/n_2)$。

前曾述及，當$\overline{x}$呈常態分配，其均數爲μ，變異數爲σ^2/n，且σ^2爲已知，則μ的95％信賴範圍爲$\overline{x} \pm 1.96\sigma/\sqrt{n}$，同樣的，若$\sigma^2$爲已知，則$\mu_1-\mu_2$的95％之信賴範圍爲$(\overline{x}_1-\overline{x}_2) \pm 1.96\sigma\sqrt{1/n_1+1/n_2}$。

一般情況下，變異數σ^2多爲未知，須先藉以求出一估計數再行估計。

估計數之一S_1^2可用下式：

$$S_1^2=\frac{\Sigma X_1^2-n_1\overline{X}_1^2}{n_1-1}$$

公式〔18－1〕

欲求出第二個估計數亦可如同上述公式求出S_2^2。再合併此二估計數成爲變異數之合併估計數(pooled estimate of the variance)，可以符號S_p^2來表示。S_p^2爲S_1^2與S_2^2的加權均數(Weighted mean)，可簡稱爲合併變異數(pooled variance)，其公式如下

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

公式〔18－2〕

若$\overline{x}_1 - \overline{x}_2$呈常態分配，該均數爲$\mu_1 - \mu_2$，變異數爲$\sigma^2(1/n_1 + 1/n_2)$，則此量數

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}}$$

呈標準常態分配。

又如果以S_p代換上式中之σ，如下：

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}}$$

便爲呈 t 分配的另一量數（quantity）。$n_1 + n_2 - 2$爲求算S_p^2的分母，也這是此 t 分配的自由度。$\mu_1 - \mu_2$的 95％信賴範圍可以S_p代換σ與 t 分配的 97.5％點，自由度$n_1 + n_2 - 2$代替 Z 分配的 97.5％點的 1.96 便可。$\mu_1 - \mu_2$的 95％信賴範圍如下：

$$(\overline{x}_1 - \overline{x}_2) \pm t[.975]\, S_p\sqrt{1/n_1 + 1/n_2}$$

$$(d.f. = n_1 + n_2 - 2)$$

ROENTGENOLOGY, LABORATORY

W-8-1 Wang : RADIOTRACER METHODOLOGY IN THE BIOLOGICAL ENVIRONMENTAL AND PHYSICAL SCIENCES $200

DENTISTRY

A-9-1 Arnold : OCCLUSAL TREATMENT — Preventive and Corrective Occlusal Adjustment, 1976 .... $140

A-9-2 A.D.A. : UPDATE IN CLINICAL DENTISTRY 1978: Selected from Oral Research Abstracts $110

B-9-1 Brewer : OVERDENTURES, 1975 .... $290

C-9-1 Colby : COLOR ATLAS OF ORAL PATHOLOGY $400

E-9-1 Eversole : CLINICAL OUTLINE OF ORAL PATHOLOGY: DIAGNOSIS AND TREATMENT, 1978 .... $320

G-9-1 Goldstein : ESTHETICS IN DENTISTRY, 1976 .... $420

G-9-2 Graber : REMOVABLE ORTHODONTIC APPLIANCES, 1977 .... $350

G-9-3 Gelb : CLINICAL MANAGEMENT OF HEAD NECK AND TMJ PAIN AND DYSFUNCTION 1977 .... $500

H-9-1 Hirshberg : PERIODONTAL PROSTHESIS, 1977 .... $270

H-9-2 Horn : PRACTICAL CONSIDERATIONS FOR SUCCESSFUL CROWN AND BRIDGE THERAPY — Biologic Considerations. Psychologic Considerations. Preventive Factors, 1976 .... $250

M-9-1 Melcher : THE SCIENTIFIC BASIS OF RECONSTRUCTIVE DENTISTRY, 1977 .... $100

P-9-1 Perel : DENTAL IMPLANTOLOGY AND PROSTHESES, 1976 .... $400

P-9-2 Prichard : THE DIAGNOSIS AND TREATMENT OF PERIODONTAL DISEASE IN GENERAL DENTAL PRACTICE $550

S-9-1 Stahl : PERIODONTAL SURGERY: — Biologic Basis and Technique, 1976 .... $360

S-9-2 Shelby : ANTERIOR RESTORATION FIXED BRIDGEWORK AND ESTHETICS, 1976 .... $360

Y-9-1 Yalisove : TELESCOPIC PROSTHETIC THERAPY — Periodontal Prosthesis Fixed and Removable $450

Z-9-1 Zerb : PROSTHODONTIC TREATMENT FOR PARTIALLY EDENTULOUS PATIENTS, 1978 .... $320

Z-9-2 Zegarelli : DIAGNOSIS OF DISEASES OF THE MOUTH AND JAWS (2nd Ed.), 1978 .... $500

醫學(包括解剖生理)

許鴻志　編　著：成人心臟病與治療……………… 200元

許鴻志　編　著：心電圖鑑別診斷……………… 300元

許鴻志　編　著：心臟病學一臨床手冊(第一冊)……………… 120元

許鴻志　編　著：心臟病學一臨床手冊(第二冊)……………… 100元

許鴻志　游天翔：心臟病學……………… 400元

蔡坤喜、[illegible] 王夫[illegible]、[illegible]：最新哮喘診斷治療……………… 70元

國　興　編　委　編：最新疾病學……………… 80元

楊亨[illegible]　編　著：實用臨床物理診斷學……………… 270元

林元[illegible] 蔡坤喜 [illegible] 合編著：最新實驗診斷解說……………… 165元

朱學琳　編　著：臨床診斷學……………… 60元

蔡坤喜　張秀平：醫師速記……………… 100元

何禮達　編　著：醫學概論……………… 80元

謝文[illegible]　編　著：性病學淺論(附個案討論)……………… 75元

[illegible]：梅氏X光診斷學(全三冊)……………… 2800元

朱學琳　編　著：實用眼科學……………… 280元

朱學琳　編　著：眼科手術……………… 280元

[illegible]：基礎心電圖學……………… 100元

許鴻志　編　譯：心電圖簡易判讀法(改訂三版)……………… 200元

醫學(包括解剖生理)

許鴻志 編 譯：中毒之診斷與治療手册……240元
國興編委會編：SIMPLE MANUAL OF MEDICINE DIAGNOSIS……160元
余幸司 編 著：由唱片學心臟的聽診……200元
國興編委會編：最新處方手册……50元
石竹根 編 著：最新處方學……250元
#國興編委會編：常用新藥內科處方學……100元
沈祖杰 編 著：彩色圖說皮膚病治療指針……700元
田明輝 編 譯：常用藥用量(小兒藥用量改訂第三版)……60元
許左爾 譯 關學純 許鴻志 監修：口腔病理(彩色圖解)……400元

余家利 編 譯：消化管內視鏡圖譜……400元
蔡瑞熊 余家利：這就是透析療法……80元
楊介人 編 譯：實用臨床內科學……330元
曾曾旂 編 著：急救與消毒法(附歷屆考題)……100元
國興編委會編譯：婦產科診療圖說……元
曾曾旂 編 著：新編公共衛生學……280元

梁維熏 編 者：最新剝脫細胞學
第一册婦科剝脫細胞學……600元
第二册呼吸系統、體腔液、剝脫細胞學、循環血液中癌細胞之研究……550元

林金權 編 著：營養與健康……55元
蕭煜倫 編 譯：彩色細胞診之基礎……190元
許鴻志 編 譯：小兒科臨床手册……180元
蕭煜倫 編 譯：簡易臨床診斷學便覽……170元
蕭煜倫 編 著：臨床常用藥品劑量手册……120元
王新榮 編 著：773彩色圖說 小兒皮膚病學(診斷與治療指針)……780元

卓倩台 傅美亮 郭滋津 編 譯：臨床診斷，處方集(第七版)……550元

許鴻志 編 譯：心臟病學之病例報告……280元
何賜元 編 譯：矯形神經學(神經高度診斷指南)……200元
劉華嚴 編 著：靜脈內營養輸液一在小兒科及小兒外科的臨床應用……65元
謝文義 編 譯：不孕症淺論……55元

潘宏照 編 著：最新腸胃放射線診斷技術……300元

陳英奉 編 譯：眼觀產科學……400元
蕭學如 許燦瑩編：產科學讀本……125元
蕭學如 許燦瑩編：兒科學讀本……125元
謝文義 編 譯：臨床小兒科學(基礎與應用)……125元
謝文義 編 譯：小兒科急救手册(診斷與處置)……85元
國興編委會編譯：皮膚病與性病學(臨床診斷與鑑別)……200元

潘宏照 編 著：性的奧密……75元
劉永馳 朱淑珍：醫藥護技術手册……120元

許世宗 譯：臨床實用呼吸系疾病之診療……元
蕭學如 譯：臨床實用皮膚科疾病之診療……85元

國興編委會編譯：臨床檢查及診斷……300元
國興編委會編譯：臨床內科診斷學……250元
蔡錫圭 等 校：解剖學實習手册……250元
李眞修 編 譯：新解剖生理學……120元

蔡長添、黃萬正
陳 瑛、劉正民：蕭氏生理學(增訂新版)……300元
林正一 合 編

醫學（包括解剖生理）

曾曾旂　編著	環境衛生重點整理題要一（含歷屆考題）	180 元
李辛然　編著	公共衛生學提要	65 元
余萬能　編著	流行病學與傳染病管理題解	70 元
曾曾旂　編著	急救與消毒（附單元複習・歷屆考題）	100 元
國興編委會編	急救法與消毒法重點提示及試題分析	50 元
余萬能　編	環境衛生學試題精解	80 元
國興編委會	公共衛生學考試複習講義	80 元

藥學

※經銷書

張文哲　編著	藥劑學要說	150 元
＃鍾振明、張育元	有機醫藥品合成化學	100 元
＃林南會　編著	維他命含量測定及鑑定	80 元
許志堯、徐型堅	藥物學要說	100 元
國興編委會編	新編藥事法規摘要（民國67年七月增訂版）	50 元
洪玉霖　陳逸光	基礎藥理學　上冊（第六版）	140 元
洪玉霖　陳逸光	基礎藥理學　下冊（第六版）	180 元
徐型堅　林建宗	藥品分析法	250 元
黃占甲　編著	藥劑學	200 元
國興編委員編	最新處方手冊	50 元
＃國興編委會編	常用新藥內科處方學	100 元
王仁澤　編著	毒物化學及實驗	250 元
吳東龍　編著	萬田氏藥理學	200 元
石竹根　編著	最新處方學	250 元
供學編委會編	常用生藥手冊	120 元
王仁澤　顏孝治	藥品定量分析及藥品鑑定各論	250 元
王仁澤　編著	藥品定性分析及一般試驗法	250 元
林金龍　編著	詳解藥理學	250 元
姜宏哲　王仁澤	藥品儀器分析	150 元
黃武雄　編著	調劑一藥品配合禁忌	200 元
許志堯　編著	有機反應合成90	60 元
許志堯　編著	有機合成演習	50 元
洪金柃　編著	實用藥理學	150 元
劉正雄　編著	新藥劑學（第三版）	400 元
顏孝治　編著	藥鑑各論與解說（附問題解答）	100 元
嘉南藥專編	最新藥劑學	80 元
林地淇　編著	藥劑學與其實驗	120 元
供學編委會編	最近藥事法規	180 元
許興智　姚秀蘭	重要藥物之作用及其化學	280 元
＃徐享堂　編著	新藥用植物學	350 元
※蔡靖彥　編著	常用藥品手冊（72年版）	345 元
※蔡靖彥　編著	常用藥的副作用與對策（1981年）	245 元
※衛生署　編	中華藥典（第三版）	800 元
黃建才　編著	葯劑學（附中華葯典第三版摘要）	270 元
呂芳雄　著	藥品的製造與衛生工程	元
黃建才　編著	葯劑學（包括調劑學）一附中華葯典第三版摘要	350 元
蕭倉農　編著	最新葯物化學（1981年）	400 元
黃建才　編著	有機葯物化學（1981年）	350 元
黃建才　編著	葯事法規（1981年）附模擬試題	220 元
曾誠齊　編著	基礎藥物化學	400 元
林宗平　編著	醫藥品化學（上冊）（下冊）	各600 元
郭桂美　著	藥物及食品管理	元
黃崇墉　著	藥理學精要	元

醫學(包括解剖生理)

黃振鐘 編	腦部CT。診斷	220元
潘宏照 編譯	胃癌與內視鏡檢查(彩色版)	600元
國興編委會編譯	臨床內科學一診斷、檢查治療	元
國興編委會編譯	新圖解醫學英語百科辭典	元
國興編委會編	圖解骨折固定保存療法	180元
國興編委會編	心臟的診斷與檢查技術	200元
余萬能 編譯	胃腸症狀之臨床用藥	170元
許文明 著	現代婦女醫學	270元
蘇新輝 著	公共衛生學	330元
李中章 編著	傳染病與流行病管制	400元
謝文義 編譯	內科疾病的皮膚症狀(彩色版)	550元
鄭懷德 著	環境衛生指引	元
國興醫學編委會編	心臟超音波的臨床診斷	320元
洪宗民 著	復健醫學概論	元
楊德明 著	急救與消毒	元
包興林 著	生物統計學概論	200元
張哲源 著	工業的安全與衛生	400元
周國群 著	作業環境的安全與衛生	400元
蘇勝章 著	內外科疾病的診察法	元
蘇勝章 著	內外科疾病的治療法	元
蘇勝章 著	內外科疾病的診斷與治療	元
國興編委會著	急診檢查手册	元
沈傳宗 著	心理學與心理衛生	元
謝文義 編譯	實用內科學(Bed-Side MEMO)	350元
楊德明 著	醫院病房的管理	元
廖春鈴 著	婦科學概論	元
吳春蓮 著	產科學概論	元
楊德明 著	家庭急救百科全書	元
徐莉美 著	女性美容與健身寶典	元
黃崇墉 著	生理學精要	元
張仁福 著	流行病學	元
張仁福 著	傳染病管理	元
李益中 著	皮膚疾病的診察與治療	元

醫師牙醫師考試用書(包括:衛生行政人員,保健員特考用書)

——請再參考基礎醫學部份——

國興編委會編	內科複習精華	85元
國興編委會編	外科複習精華	80元
國興編委會編	婦產科複習精華(增訂二版)	70元
國興編委會編	小兒科複習精華	45元
林大春 編著	國家醫師檢覈考試總整理——臨床醫學	110元
林大春 編著	國家醫師檢覈考試總整理——基礎醫學	75元
徐志淵 編著	公共衛生複習精華(增訂版)	65元
鄧榮坤 編著	公共衛生總複習(附三民主義精選考題)	65元
謝文義 編著	公共衛生條文式精選複習題解	55元
張耀琦 編著	公共衛生保健員模擬試題	110元
張耀琦 編著	醫事人員特考模擬試題	150元
謝文義 編著	環境衛生學與微生物學條文式精選題解	75元
張耀琦 編著	口腔(牙醫)衛生人員模擬試題	70元
國興編委會編	公共衛生複習題解	75元
曾啓旂 編著	人類傳染病學複習指南一(含歷屆高,普特考試題	180元

檢驗學（包括：免疫學、生化學、病理學）

杜啓安	編　譯	臨床生化基論	450元
張來發	編　著	臨床生化診斷學	85元
邱金賜	編　著	病理切片技術與組織染色法	250元
杜啓安	編　譯	病理解剖技術學	70元
林元龍	張秀平	常用試劑手冊	120元
薛樹清	編　譯	基礎免疫學	70元
楊炳炘	編　譯	免疫學要義	125元
※游世規	編　著	臨床細菌學	240元
蔡喜坤 林元龍	合　編	最新實驗診斷解說	165元
蔡坤喜	蔡志堯	生化學原理	80元
葉東柏	編　著	生物化學指南	250元
何禮達	編　著	臨床生理檢驗學	60元
U. S. A.	編	各類致病性微生物彩色標本圖表	80元
U. S. A.	編	血球發展型態及各型惡性貧血，白血病之彩色標本	80元
國興編委會	編	實用血庫作業與病理檢驗技術	50元
何禮達	編　著	臨床寄生蟲學（彩色版）	400元
何禮達	編　著	微生物學（彩色版）	550元
陳辰權	編　譯	實用臨床化學須知	250元
何禮達	戴金蓮	臨床實驗診斷儀器手冊	160元
陳辰權	編　著	生物化學實驗	220元
杜啓安	編　譯	血液學技術（Technical Hematology）	280元
陳英華	編　著	病理學常識	元
杜啓安	編　譯	免疫學（roitt，第四版）	150元
何禮達	編　著	臨床檢驗學（彩色版）	280元
邱春風	李達仁	臨床微生物學（上册）	110元
邱春風	羅光榮	臨床微生物學（下册）	160元
戴照隆	邱春風	臨床生化學（上）	200元
戴照隆	邱春風	臨床生化學（下）	180元
張文瑞	編　著	臨床血清學	140元
陳辰權	編　著	生化儀器分析	元
杜啓安	編　譯	組織病理技術手册	200元
楊介人	編　譯	病理學討論	115元
高源彬	編　著	臨床鏡檢學圖鑑（彩色）（尿沈渣，血液像骨髓像，寄生蟲微生物）	700元
高源彬	編　著	尿沈查鏡檢圖鑑（彩色）	200元
高源彬	編　著	微生物鏡檢圖鑑（彩色）	250元
高源彬	編　著	寄生蟲鏡檢圖鑑（彩色）	130元
高源彬	編　著	血液像與骨髓像鏡檢圖鑑（彩色）	200元
謝慶良	編　著	臨床免疫與技術學	170元
羅光榮	編　著	臨床生化學及螢光免疫分析檢驗學	180元
國興編委會	編	臨床檢查知識總整理　血液學	90元
國興編委會	編	臨床檢查知識總整理　生化學	70元
黃耀宜、柯珍珍		臨床生化檢驗技術手册	250元
許俊男	編　著	有機生化基礎	100元
國興醫學編委會	編	臨床細菌培養成分便覽	元
黃崇墉	著	病理學精要	元

醫檢師考試用書（請再參考基礎醫學）

國興編委會	編	最新醫檢師檢覈考試複習精華（增訂版）	200元
鄧榮坤	編　著	最新醫檢師檢覈全集	85元

新編生物統計學

中華民國七十三年 四 月初版

編著者：包 興 林
發行人：楊 國 藩
發行所：國 興 出 版 號
新竹市西門街 262 號
電話：(035) 223129 號
行政院新聞局局版台業字第 0465 號
總經銷：黎 明 書 店
新竹市中正路 72 號
電話：(035) 229418 號
郵政劃撥七〇二七三號

實價： 200 元